AD-A118 429 AIR FORCE ENVIRONMENTAL TECHNICAL APPLICATIONS CENTER--ETC F/G 9/2
OBJECTIVE ANALYSIS OF CLIMATOLOGICAL PROBABILITY DATA.(U)
JUL 82 B E LILIUS, F C WIRSING, R M COX
UNCLASSIFIED USAFETAC/TN-82/003 SBI-AD-E850 181 NL

USAFETAC/TN-82/003

OBJECTIVE ANALYSIS OF CLIMATOLOGICAL PROBABILITY DATA

BY

Major Bryan E. Lilius

1Lt Frederick C. Wirsing

2Lt Robert M. Cox

JULY 1982

UNITED STATES AIR FORCE
AIR WEATHER SERVICE (MAC)
USAF
ENVIRONMENTAL
TECHNICAL APPLICATIONS
CENTER
SCOTT AIR FORCE BASE, ILLINOIS 62225

REVIEW AND APPROVAL STATEMENT

USAFETAC/TN-82/003, Objective Analysis of Climatological Probability Data, July 1982, is approved for public release. There is no objection to unlimited distribution of this document to the public at large, or by the Defense Technical Information Center (DTIC) to the National Technical Information Service (NTIS).

This technical publication has been reviewed and is approved for publication.

DR. PATRICK J. BREITLING
Chief Scientist

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: USAFETAC/TN-82/003	2. GOVT ACCESSION NO.: AD-A118 429
	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): OBJECTIVE ANALYSIS OF CLIMATOLOGICAL PROBABILITY DATA	5. TYPE OF REPORT & PERIOD COVERED: Final January 1981-March 1982
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Maj Bryan E. Lilius; 1Lt Frederick C. Wirsing; 2Lt Robert M. Cox	8. CONTRACT OR GRANT NUMBER(s):
9. PERFORMING ORGANIZATION NAME AND ADDRESS: USAF Environmental Technical Applications Center/DND, Scott AFB, Illinois 62225	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: USAF Environmental Technical Applications Center, Scott AFB, Illinois 62225	12. REPORT DATE: July 1982
	13. NUMBER OF PAGES: 105
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Report of work done under USAFETAC Project 2502. Field 4

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Objective Analysis
Data Compaction
Climate Spreading
Ceiling
Visibility
Germany Climate

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes a method for compacting, accessing, and objectively analyzing ceiling/visibility probability data. Unconditional cumulative probabilities over the southern half of West Germany are analyzed. The technique used to compact the data uses less than one-eighth of the computer storage normally required. The accuracy of the data was only slightly impaired. Three objective analysis algorithms were investigated; Barnes, Janota, and nearest neighbor. The Barnes method performed best. Independent data were estimated, using this technique with a mean error of 3.2 percent.

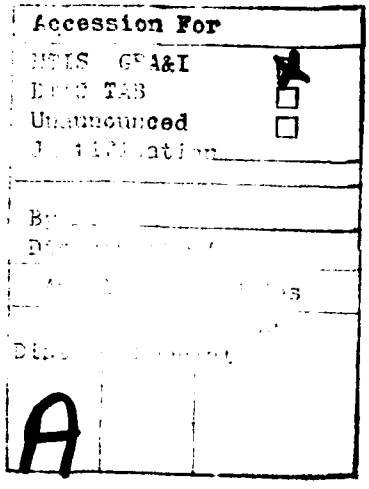
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PREFACE

This report is a discussion of the work performed in conjunction with USAFETAC Project 2502. The project was requested by Air Weather Service (AWS) through Air Force Global Weather Central (AFGWC) in response to an AWS validated requirement that the USAF Environmental Technical Applications Center (USAFETAC) provide climatological probabilities at locations with no historical observational record. This requirement was formally identified and documented in a survey of Worldwide Military Command and Control System (WWMCCS) staff weather officer users. However, development and implementation of this capability will significantly add to USAFETAC's ability to support all customers. The specific purpose of Project 2502 is to provide a demonstration and evaluation of the technique proposed by USAFETAC Data Base Development Section to meet this requirement. The results documented in this report are generally favorable.

We gratefully acknowledge the numerous suggestions and contributions of Majors Pershing Hicks and Albert Boehm, and Captain Emil Berecek III.



CONTENTS

ILLUSTRATIONS

Estimated Probability Fields

Estimated Probability Fields

TABLE

Chapter 1

INTRODUCTION

1.1 Historical Background

Since the early 1970's, the US Air Force Environmental Technical Applications Center (USAFETAC) has wrestled with the problem of how best to meet the climatological needs of the Worldwide Military Command and Control System (WWMCCS) community. Until recently, the climatological information requirements of WWMCCS users were not well defined, although these requirements were fairly accurately perceived to be a rapid response providing typical USAFETAC products, often for locations or areas which lacked a good historical record of weather observations.

Such requirements were generally addressed by USAFETAC WWMCCS planners with an insistence upon acquiring a very large and extremely expensive mass storage device. On such a mass storage device, a significant portion of our USAFETAC data base could be stored in a rapid access mode, and general applications programs then run against the data in a manner similar to our standard operations which access data from magnetic tape. Past proposals such as this did not address the problem of data for nonobservation locations. The net result of these plans has chiefly been to identify the prohibitive cost of such a change in our mode of operation.

In the fall of 1979, the USAFETAC Data Base Development Section (DND) distributed a survey to all WWMCCS users requesting them to identify their specific climatological information requirements. These survey responses showed that the primary requirements were for the same sort of products USAFETAC routinely provides its customers. The only stated requirements somewhat peculiar to the WWMCCS were an occasional, but urgent, need to obtain this information within 1-2 hours and a need to obtain estimates of climatological probabilities at locations with no historical observational record.

In addressing these requirements, DND recognized there were numerous possible ways to attack these problems. In a September 1980 USAFETAC report to Air Force Global Weather Central (AFGWC) summarizing the results of the customer surveys four possible responses were considered. These included:

a. Decide that the problem was not likely to be solved at reasonable cost and thus doing nothing.

b. Build a summarized data set which would be published and disseminated prior to its required use. This data set would have to be inclusive and specific enough to satisfy nearly all stated requirements.

c. Obtain a large mass storage device.

d. And, finally, incorporate developing capabilities within USAFETAC into a technique which would begin to correct some of the shortfalls in USAFETAC capabilities.

The advantages and disadvantages of each of these approaches were discussed in the report, and the fourth alternative was recommended.

In the fall of 1980 USAFETAC requested through AFGWC that Air Weather Service (AWS) validate the two requirements which USAFETAC was unable to satisfy. AWS validated the requirement to provide climatological information for locations and areas which had no observations. However, they did not validate the requirement to respond within 1-2 hours, reserving that decision for a later date. DND proposed that a limited demonstration of a technique which would estimate climatic information between stations be provided. This technique, representing a potential USAFETAC capability, would incorporate techniques previously developed within the USAFETAC Aerospace Sciences Branch (DN) for use by other agencies. This report is a description of the results of this technique development.

1.2 Description of Problem

1.2.1 Estimating Climate Data Between Stations. It is well-known that climatological information is primarily a summary of observed weather at various locations. Meteorological observation sites are scattered about the globe in irregular networks. Some areas have dense networks while others are sparse. Indeed, large areas of our globe have practically no observation sites at all.

It is by no means clear what is the best method of estimating climatological information for these points and areas which have no observational data.

For synoptic data used in day-to-day forecasting, the method used is simply interpolation between actual observation sites. For many continuous variables such as pressure and temperature, this method for filling in the holes is the best one. Local terrain features and other geographic influences cause other types of weather variables, such as cloud ceiling, cloud cover, visibility, precipitation, and surface wind to vary considerably and on a scale much smaller than the distance between observation sites (even in very dense networks). What is the best way to account for this small-scale variability in these parameters?

The problem is to estimate the climatological or empirical probability of occurrence of operationally important weather parameters at locations which have no observational record. The necessity to accomplish this estimate with an accuracy that is operationally useful is included, i.e., operationally useful for

military operations such as launching and recovering aircraft, air strikes, paradrops, surface movements of combat troops, etc.

One way to attack the problem is to interpolate between the climatological probabilities observed at surrounding stations. This method ignores (or accepts) the inaccuracies due to smaller scale variability in the data. It detects and displays the variability which occurs on a scale as large or larger than the distance between observation sites.

Another method is to relate the observed climatological distributions to the surrounding terrain features and other identifiable geographical influences. Then, knowing this relationship at a location, estimate the climatological probability of the weather event of interest. This method is appealing for a number of reasons. It attempts to consider the small-scale variation which is known to occur. It should then be possible to distinguish the mountain station from the valley station, coastal from continental, and so on.

This second method also has its drawbacks though, two of which need to be mentioned in this discussion:

a. An intimate knowledge of the geography of a requested location must be known a priori.

b. This relationship is usually determined by linear regression and yet the relationship is a nonlinear one with interaction between the synoptic and small-scale flow and the geographical features. It is exceedingly difficult to capture and describe this relationship very well at reasonable cost.

We attempted to optimize a technique to interpolate between stations in a data rich region. Enough time was allowed to do a thorough evaluation of this technique. It is still, perhaps, an open question whether the accuracy obtained is operationally useful. However, the accuracy obtained from the techniques developed or modified for this project exceeds any previous attempt of which we are aware.

1.2.2 Synoptically Formatted Data Set. In order to interpolate any kind of data between observation sites the data must be available simultaneously for all of the sites at the time of interest. For summarized climate data, this is not synoptic in the usual sense, but it is similar. Climate data implies statistical information of some kind, data averaged or counted over a specific period. For example, if one is concerned about morning ceilings in June, then data is needed which has been selectively averaged or counted over several years for that month and time of day for all the stations surrounding the location of interest. It is in this sense that a synoptically formatted data set is discussed.

For this project we chose to work with cloud ceiling and visibility data for three reasons. First, these are parameters of great interest to the military community. Second, these are parameters which are greatly affected by local terrain influences. Third, a ceiling and visibility data set suitable for this project was already available within USAFETAC/DN. This data set consisted of ceiling and visibility ogives for all western Europe and had been compiled as a part of a different project. These ogives are cumulative distribution functions (CDF) which give the probability that the ceiling or visibility will exceed a specified threshold. These CDFs or ogives are actually discrete values, consisting of probabilities for 32 standard RUSSWO (Revised Uniform Summary of Surface Weather Observations) categories for ceiling and 15 for visibility. The CDF data are unconditional in that they are not conditioned on any event other than time.

From this data set all of the available suitable data for the southern half of West Germany was extracted. There were 81 stations for the area shown in Figure A.1 of the Appendix, and are listed in Table A.1 along with their WMO (World Meteorological Organization) numbers and elevations. This is an area with very dense data coverage. It is also an area with highly variable terrain features. Both of these aspects of the data set were important to the project. With dense data coverage, it is possible to experiment with data density by withholding some of the available data. This, of course, is not possible if the actual data coverage is sparse. The variable terrain makes it possible to study the effects of terrain on the techniques employed.

1.2.3 Compacted Data Set. While this CDF data for these 81 stations is not a large amount of data, a generalized method using CDF data for the whole world requires a tremendous amount of such data. This would certainly be more data than would fit on any near-term configuration of USAFETAC random access devices. Therefore, it is necessary to either reduce the storage required for the data in some way or buy more disks. The first alternative is of course, preferable. The problem, then, is to reduce the storage space required for this CDF data while minimizing the loss in recoverable accuracy resulting from this reduction. This reduction process is called "compacting" the data. Chapters 2 and 3 of this report describe the method we employed to compact the data and Chapter 6 describes the statistical accuracy of the technique. The development of this technique was one of the most important parts of this project.

A typical CDF is a summary of perhaps 500 ceiling or visibility observations into 32 or 15 numbers, respectively. These are the discrete values of the probability that the ceiling or visibility will exceed these standard threshold values, such as found in a RUSSWO. These numbers then require only 3-6 percent of the space required to store the observations themselves. The compaction technique that was developed reduces the storage required still further, requiring three words of storage for ceiling and two for visibility for each CDF stored. This is now only 10-13 percent of the space required to store the actual CDF data

and is less than 1 percent of the space required to store the original observation data.

Once these data are compacted and synoptically formatted, there remains only the problem of accessing the data and analyzing them objectively for the area of interest. Chapter 5 describes three analysis techniques we investigated and Chapter 6 describes the statistical results of each of these techniques.

The program "WWMX," which is the heart of this project, is the tool which processes a user request by accessing the data, performing the objective analysis, and providing the desired answer using any one of many possible output options. This program is described more completely in Chapter 4.

The project objective was to develop and evaluate thoroughly an interpolation model that could be used to obtain estimates of climatic data at locations between stations. No claim is made that the techniques developed have been optimized. However, the results described in the following chapters are certainly promising. Serious consideration should be given for further development and implementation of these techniques as an operational capability at USAFETAC. Chapter 7 describes the direction such development and implementation should take and attempts to estimate the costs involved.

CHAPTER 2

FITTING THE DATA

Figure 1 depicts an ideal cumulative distribution (of frequency) function (CDF), which is a continuous function of the threshold parameter. To approximate the curve for a real weather parameter such as ceiling or visibility, raw data (observations) can be tallied or "bean counted" to determine the percentage frequency of occurrence of the parameter above certain thresholds. The program DNOEDCV was used to perform this task using a DATSAV POR for input and standard RUSSWO categories for thresholds. The result is a CDF described by a number of discrete points. This set of discrete points is called an ogive.

Traditionally, in an attempt to compact the CDF, considerable effort has been invested in determining what were suitable forms of the function. Functions with only a few coefficients were preferred, because, by retaining the form of the function and a few coefficients (determined by numerical fitting of the function to the RUSSWO probabilities), the CDF could be "recovered." Furthermore, because of the nature of a continuous function, all one had to do to obtain the percen age frequency of occurrence was to supply the desired threshold and evaluate the function. Other characteristics of specific functions make them attractive for certain purposes. The original project plan called for the use of such a method which was already in production use. Examination of the statistics of "goodness of fit" revealed that these functions (log cubic for ceiling and inverse linear for visibility) fit the data with root-mean-square (RMS) errors in the vicinity of 3 percent and maximum errors of 10 percent. This seemed unacceptable. If errors of this magnitude existed before any objective geographic analysis took place the final result seemed almost sure to be less than satisfactory.

Several other techniques for fitting CDFs that were available at USAFETAC were investigated, including cubic splines, Weibull and Burr curves. Another possibility which occurred to us was to select certain points and connect them by straight lines to approximate the RUSSWO unconditional probabilities. This latter technique proved to be superior both in "goodness of fit," in ability to compact the CDF, and in processing time. This technique, which is termed line-segment-selection (LSS), and will be described in more detail. Also, it has the advantage of being general; i.e., it will work for any parameter that can be described by a CDF.

Figure 2 depicts an ideal CDF, described by ten hypothetical RUSSWO threshold probabilities and by line-segment-selection. The LSS CDF in this case consists of four line-segments described by five points, LSS_1 through LSS_5. Obviously, the problem is to select five points which reasonably describe (with low error)

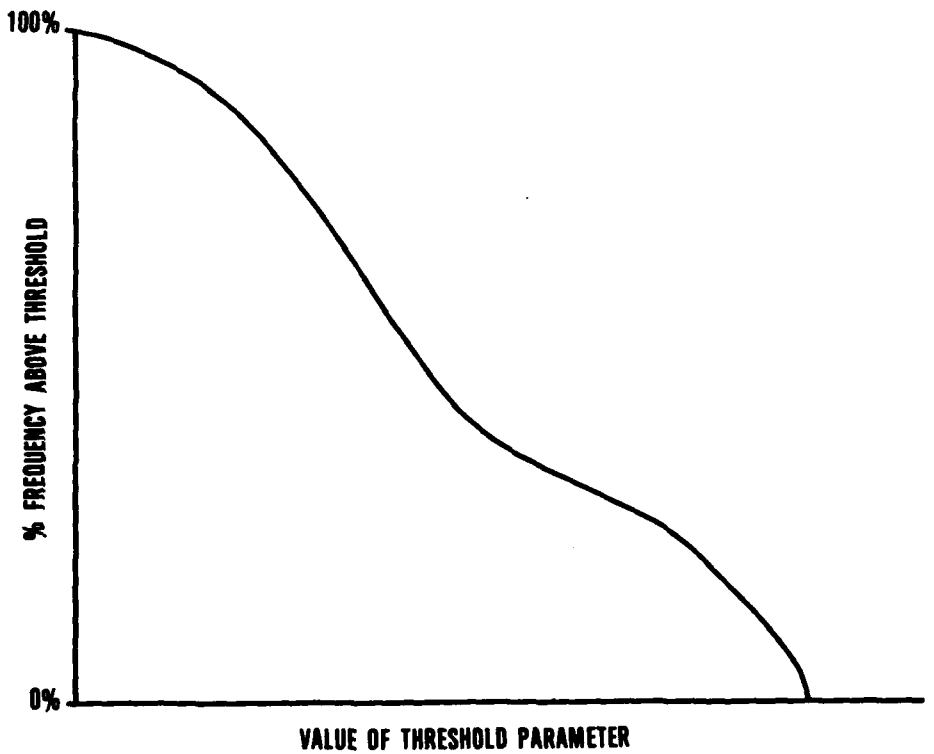


Figure 1. Cumulative Distribution Function.

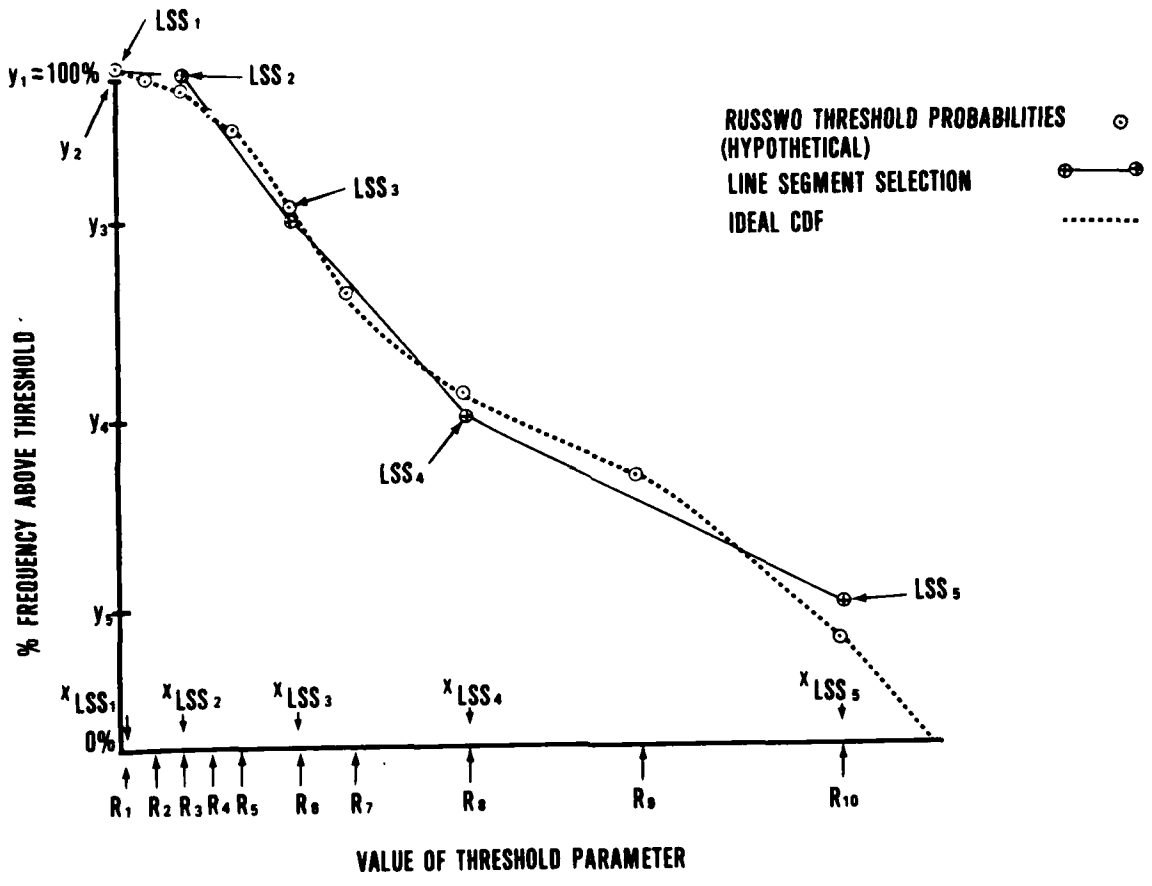


Figure 2. Description of Cumulative Distribution Function by RUSSWO Threshold Probabilities and Line Segment Selection.

the ideal CDF. Note that 10 numbers (x and y for each of five points) are required to describe this four-segment CDF. The RUSSWO CDF also requires only 10 numbers to describe it because the x-values (thresholds) are known a priori, i.e., they are standard thresholds. It appears that nothing is to be gained by using the LSS CDF. However, if the same "standardization" is applied to the LSS CDF as is applied to the RUSSWO CDF, the LSS CDF can be described with five numbers. The "standardization," in this case, amounts to setting ${}^{x}LSS_1 = R_1$, ${}^{x}LSS_2 = R_3$, ${}^{x}LSS_3 = R_6$, ${}^{x}LSS_4 = R_8$, and ${}^{x}LSS_5 = R_{10}$. Now only five y-values need to be determined and stored to recover this LSS CDF. When applied to a "fitting" technique this means that the x-values are predetermined. Only the values of y need to be determined. Also note that the ideal CDF always passes through the point (0,100 percent). This fact can be "built in" to fitting, storage, and retrieval routines to avoid the necessity of storing the y-value (100) of LSS_1. Only four y-values need to be found and stored.

The next step is to devise a numerical fitting routine to generate values of y that would result in a LSS CDF with low RMS and maximum errors. Devising an analytic set of equations which would incorporate the constraints of one segment passing through (0,100), fixed or "standardized" x-values, and monotonically decreasing y-values, to be solved by a numerical least squares method was, at least temporarily, out of reach of these analysts. So a fitting scheme (LSS) was devised which admittedly does not necessarily converge on the optimum values of y (in the least squares sense), but does result in a fit whose statistics of goodness of fit are excellent, indeed they are superior to those of any other fitting schemes investigated.

Consider the 10 category RUSSWO CDF described in Figure 3. If the five standard LSS CDF values are assigned a priori to R_1, R_3, R_6, R_8, and R_{10}, as shown, then the problem is to determine y_2 through y_5. The problem was approached as follows.

Consider the three data points ${}^{y}R_1$, ${}^{y}R_2$, and ${}^{y}R_3$ which will be approximated by the first selected segment $\overline{LSS_1\ LSS_2}$. If the method of least squares were used to fit a straight line through these three points, a set of two simultaneous equations, called normal equations, would have to be solved to determine the equation of the line. Remember that it is necessary to constrain the first segment to pass through (0,100), therefore, the y-intercept is known. Only the slope needs to be determined. A single normal equation of the form;

$$m = \frac{\Sigma x_x - b\Sigma x_k}{\Sigma x_k^2} \qquad (1)$$

where b is known, needs to be solved. If this normal equation is applied to the first three data points to determine the slope (m) and the y-intercept is used as a known point, y_2 can be determined by

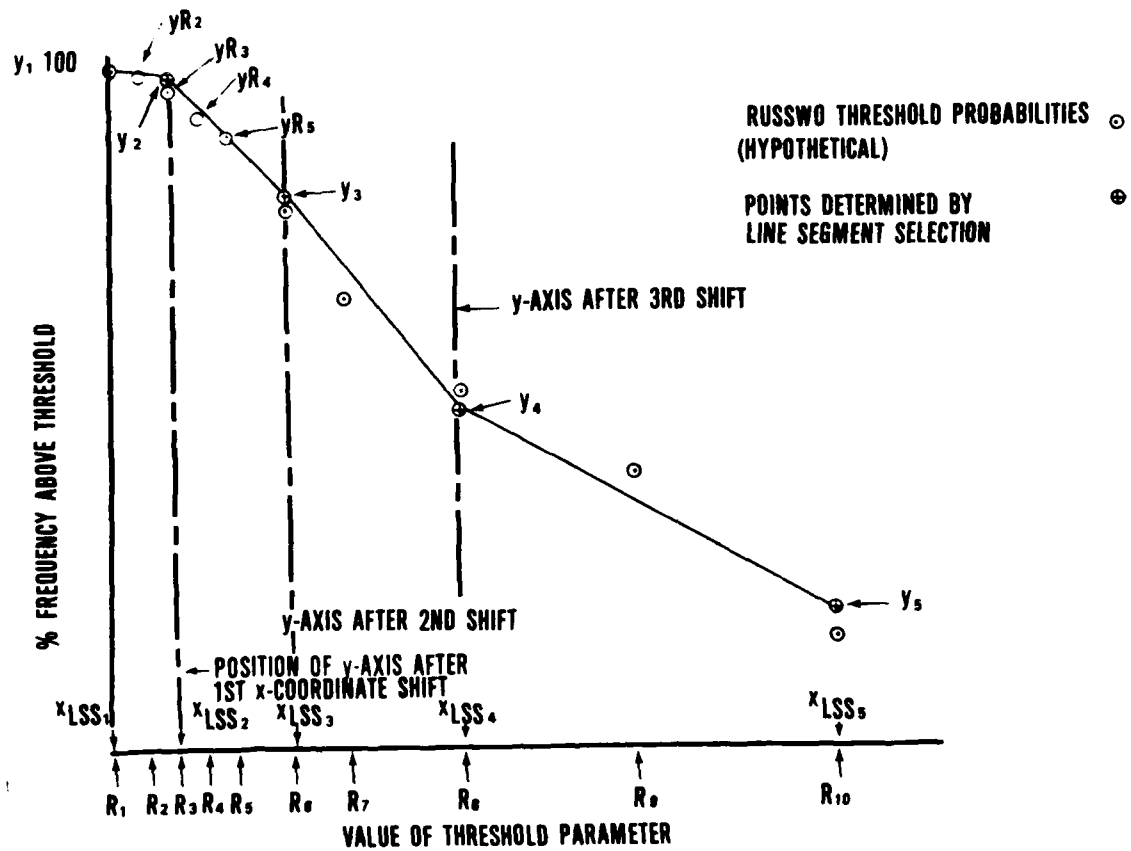


Figure 3. Description of Line Segment Selection Process (by a single series of forward regressions).

$$m = \frac{y_2 - y_1}{x_{LSS_2} - x_{LSS_1}} \tag{2}$$

or

$$y_2 = m(x_{LSS_2} - x_{LSS_1}) + y_1 \tag{3}$$

where m = slope
x_{LSS_1} = value of threshold parameter
x_{LSS_2} = value of threshold parameter
y_1 = y-intercept = 100

The best (least squares) line-segment constrained to pass through (0,100) describing the first three data points has been determined.

The second segment, $\overline{LSS_2\ LSS_3}$, can be determined in a similar manner only if an x-coordinate shift is performed as shown in Figure 3. To perform the x-coordinate shift the x-value at R_3 is subtracted from the x-values at R_3 through R_{10}; thus, the x-value at R_3 is 0. This allows the newly found y_2 to be the y-intercept. Using the single normal Equation (1) again on the data points y_{R_3} through y_{R_5} will yield the slope of a line constrained to pass through y_2. Then y_3 can be determined from Equation (3). Repetition of this coordinate shift and application of the normal equation on the intervening data points will yield y_4 and y_5, and a four segment LSS CDF will have been determined. Such a method is extremely fast and attractive in its simplicity and indeed the statistics of fit are outstanding. It must be emphasized that this method <u>does not</u> produce the <u>least possible</u> errors of fit given the constraints already stated. It is clear that y_2 is strictly determined by the first three data points. Yet the second segment which is to describe y_{R_3}, y_{R_4}, and y_{R_5} is constrained to pass through (x_{LSS_2}, y_2). It can be seen that succeeding line segments after the first <u>are not</u> necessarily the best fit through their respective intervening data points.

In order to reduce this problem, a refinement was attempted (see Figure 4). Using the previously described method y_2', y_3', and y_4' were determined in what was termed a "forward pass". Then an x-coordinate shift was performed so that the y-axis passed through R_{10} and the regression was performed by applying the normal equation to $y_{R_{10}}$, y_{R_9}, and y_{R_8}, and determining y_4''. The x-coordinate shift was repeated and, similarily, y_3'' and y_2'', were obtained in a "backward pass."

The values of y_2' and y_2'' were each weighted or "averaged" in an arbitrary way to yield y_2. The value of y_4 was otained similiarly. To determine y_3, a forward pass was performed beginning at y_2 to determine y_3' and a backward pass beginning at y_4 to obtain y_3''. Again, y_3' and y_3'' were averaged to obtain y_3. Use of this refinement further improved the statistics of fit. However, this technique does not ensure that the resultant LSS CDF is monotonically decreasing. Therefore a

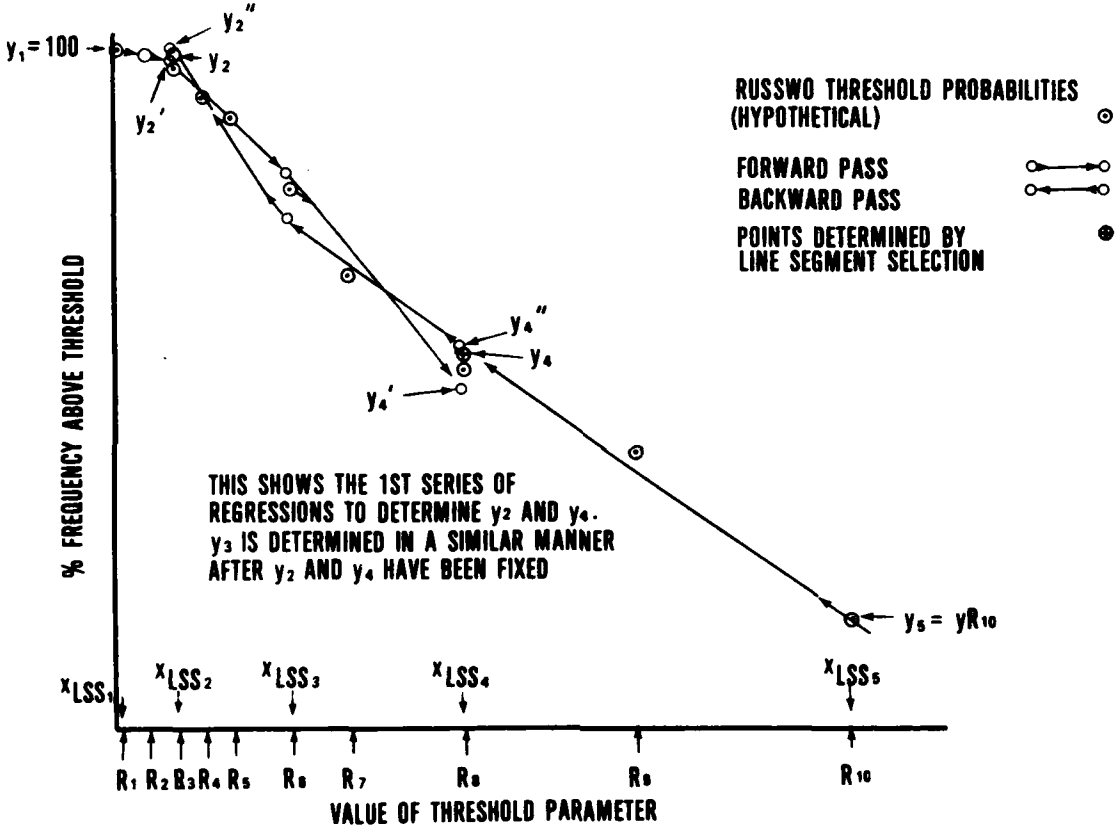


Figure 4. Description of Line Segment Selection Process (by forward-backward regression).

crude "filter" was added and the values of y were passed through the filter. For example, if y_4 were less than y_5, y_4 is forced to be equal to y_5. Then y_3 is compared to y_4 and so on until all values of y have been checked. Numerous fits were run to test this technique and the results, again, were outstanding. However, in a significant number of cases the errors of fit exceeded what was "normal" for this technique. It was noted that in these cases, if the actual data points corresponding to ${}^{x}LSS_1$ through ${}^{x}LSS_5$ were used as y-values the statistics of fit improved. This type of line segment selection is depicted in Figure 5.

Thus, the method of line-segment-selection became two-fold: (1) Perform LSS by forward-backward regression and compute RMS and maximum errors. (2) Perform LSS by picking data points and compute RMS and maximum errors. Objectively, by weighting the RMS and maximum errors, decide which set of y-values to store. It should be noted that in these fitting routines, when a y-value was computed it was immediately "rounded" to the nearest value that could be stored in the compaction scheme to be described later. The "rounded" y-values were used in all evaluations of statistics of fit, since it would be these "rounded" values that would eventually be recovered from storage to describe the RUSSWO CDF being fit.

This two-fold approach to line-segment-selection is what was used to fit and compact ceiling and visibility RUSSWO CDFs. Details of the technique applied are as follows:

CEILING

1. Twelve line-segments were used to describe the CDF.

2. The 12 y-values were "stored" in three 32-bit words.

3. The "standard" LSS categories used were:

$${}^{x}LSS_1 = RUSSWO_1 = 0 \text{ ft}$$
$${}^{x}LSS_2 = RUSSWO_2 = 100 \text{ ft}$$
$${}^{x}LSS_3 = RUSSWO_4 = 300 \text{ ft}$$
$${}^{x}LSS_4 = RUSSWO_6 = 500 \text{ ft}$$
$${}^{x}LSS_5 = RUSSWO_8 = 700 \text{ ft}$$
$${}^{x}LSS_6 = RUSSWO_{11} = 1000 \text{ ft}$$
$${}^{x}LSS_7 = RUSSWO_{13} = 1500 \text{ ft}$$
$${}^{x}LSS_8 = RUSSWO_{16} = 2500 \text{ ft}$$
$${}^{x}LSS_9 = RUSSWO_{18} = 3500 \text{ ft}$$
$${}^{x}LSS_{10} = RUSSWO_{20} = 4500 \text{ ft}$$
$${}^{x}LSS_{11} = RUSSWO_{22} = 6000 \text{ ft}$$
$${}^{x}LSS_{12} = RUSSWO_{24} = 8000 \text{ ft}$$
$${}^{x}LSS_{13} = RUSSWO_{26} = 10{,}000 \text{ ft}$$

4. Six series of forward-backward regressions were performed. The points determined in each series and the weights applied to y' and y'' were:

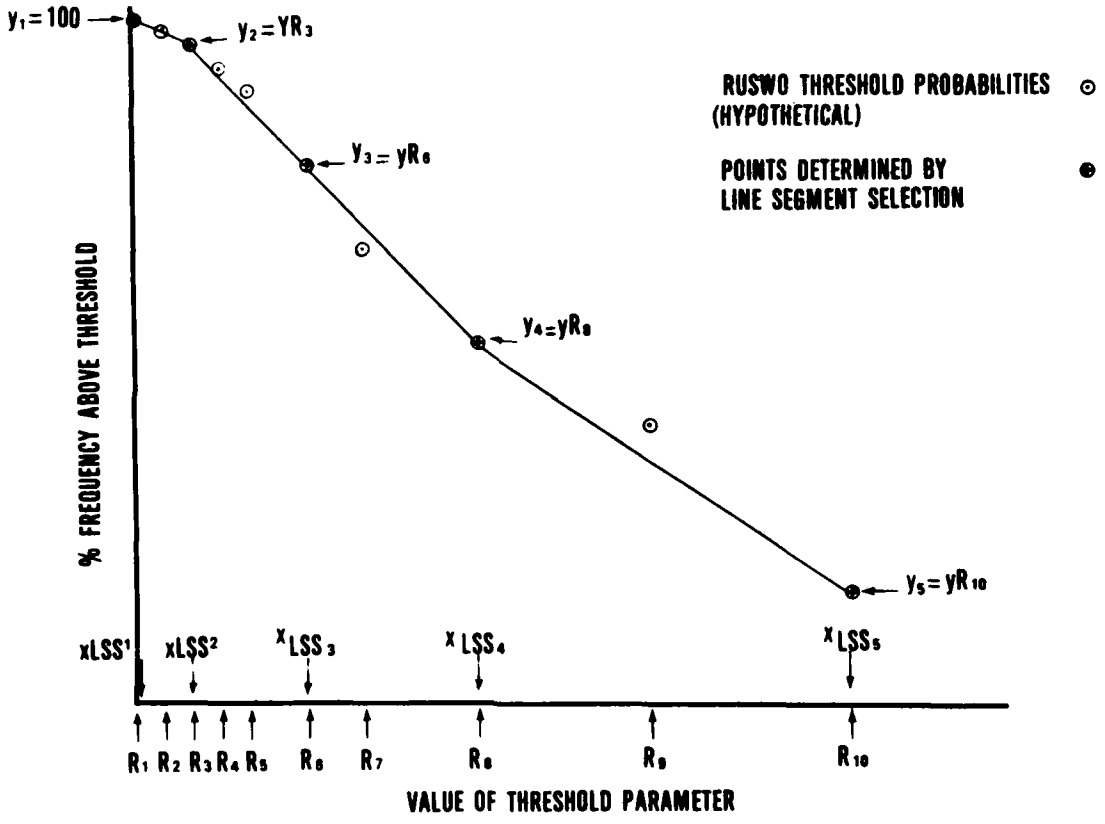


Figure 5. Description of Line Segment Selection Process (by connecting selected data points).

1st series	$y_1 = 100$		$y_{13} = {}^{y}R_{13}$
	$y_2 = 0.9y_2' + 0.1y_2''$	and	$y_{12} = 0.9y_{12}'' + 0.1y_{12}'$
2nd series	$y_3 = 0.8y_3' + 0.2y_3''$	and	$y_{11} = 0.8y_{11}'' + 0.2y_{11}'$
3rd series	$y_4 = 0.8y_4' + 0.2y_4''$	and	$y_{10} = 0.8y_{10}'' + 0.2y_{10}'$
4th series	$y_5 = 0.7y_5' + 0.3y_5''$	and	$y_9 = 0.7y_9'' + 0.3y_9'$
5th series	$y_6 = 0.6y_6' + 0.4y_6''$	and	$y_8 = 0.6y_8'' + 0.4y_8'$
6th series	$y_7 = 0.5y_7' + 0.5y_7''$		

VISIBILITY

1. Eight line-segments were used to describe the CDF.

2. The 8 y-values were "stored" in two 32-bit words.

3. The "standard" LSS categories used were

$x_{LSS_1} = RUSSWO_1 = 0$ meters = 0 miles
$x_{LSS_2} = RUSSWO_3 = 498$ meters = 1/4 miles
$x_{LSS_3} = RUSSWO_6 = 1206$ meters = 3/4 miles
$x_{LSS_4} = RUSSWO_8 = 2011$ meters = 1-1/4 miles
$x_{LSS_5} = RUSSWO_{10} = 3218$ meters = 2 miles
$x_{LSS_6} = RUSSWO_{12} = 4827$ meters = 3 miles
$x_{LSS_7} = RUSSWO_{13} = 6436$ meters = 4 miles
$x_{LSS_8} = RUSSWO_{14} = 8045$ meters = 5 miles
$x_{LSS_9} = RUSSWO_{15} = 9654$ meters = 6 miles

Two series of forward-backward regressions were performed. The points determined in each series and the weights applied to y's and y''s are

1st series	$y_1 = 100,$	$y_2 = 0.8y_2' + 0.2y_2'$ and	$y_5 = 0.8y_5'' + 0.2y_5'$
2nd series	$y_3 = 0.7y_3' + 0.3y_3''$	and	$y_4 = 0.7y_4'' + 0.3y_4'$

Chapter 3

DATA COMPACTION AND STORAGE

3.1 Compaction of Decimal Probabilities Into a Single 32-Bit IBM Word

One must consider the 32-bit word used by the USAFETAC IBM 4341 which would be the computer used if this scheme were implemented at USAFETAC. Suppose one desired to store four y-values in a single 32-bit word. Eight adjacent bits could be allocated to each y-value, provided each y-value could be expressed as an integer. This is not too difficult; simply assume the decimal point in, say, 98.6, to give 986. An actual decimal value could be provided by retrieval software. A single 8-bit segment can store only an integer between 0 and 2^8-1, or 256. Realize that a y-value ranges from 0 percent to 100 percent. If one simply attempts to store the actual decimal value (with an assumed decimal point) then y could be expressed only to the nearest percent. However, if one lets 0 percent be represented by 0 and 100 percent represented by 250, in effect the 0-100 percent range has been scaled to the integers 0-250. The "scaling" allows an integer between 0 and 250 to represent a percentage value between 0 percent and 100 percent to the nearest 0.2 of a percent. This scaling from 0-250 can be accomplished on the three right-most 8-bit segments of the 32-bit IBM word. However, the left-most bit in the word is the sign-bit, so the largest integer that can be stored in the left-most 8-bit segment is $+2^7-1 = +128$ and the smallest integer is $-2^7-1 = -128$. So, the y-value to be stored in the left-most 8-bit segment must be scaled, not from 0-250 but from -125 to +125. The same 0.2-percent resolution is obtained. Finally, multiply the left-most scaled y-value by 2^{24}, and (proceeding to the right) multiply the remaining three scaled y-values by 2^{16}, 2^8, and 2^0, respectively. The sum of these four results is a "signed" 32-bit binary number that represents four y-values to the nearest 0.2 percent of the "fitted" values of y.

3.2 Storage of Words in Files

Two unformatted random-access files were created to store the compacted words describing the LSS CDFs. One file stores the ceiling CDFs and the other stores the visibility CDFs. The first record in each file contains information such as the number of stations in the file, number of words needed to store a single CDF, the values of the RUSSWO thresholds and other extraneous information required for objective analysis. All succeeding records in each file have a record-length (in words) equal to the number of stations in the file. For example record two contains the 81 WMO numbers of the stations in the file. There are a total of four such "overhead" records in each file. The remainder of the file contains the compacted words describing the 96 LSS CDFs for each station in the file. It should be mentioned that the "standard" LSS thresholds used for each station are

also stored. As was described earlier, the LSS thresholds were set before fitting, and all stations fitted had the same LSS thresholds. However, it was recognized that the selection of standard thresholds specific to a particular station would improve the fit, therefore, the ability to store the individual station's standard LSS categories was incorporated in the file structure. For the 81 stations in the ceiling file 23,571 words are required to store the 7776 CDFs. Likewise, for the 81 stations in the visibility file 15,714 words are required.

CHAPTER 4

PROGRAM "WWMX"

Access to station data files is provided by the program "WWMX." This program has five functions.

a. Read the user request. The user must supply the desired weather parameter, thresholds, months, times, locations, and type of probability (unconditional or joint). Even large-scale requests can be entered easily; the need for repetitious entries is virtually nonexistent. One can also direct output to disk file and retain sets of locations for use in subsequent runs.

b. Access appropriate station data file(s). Access is random. Words for all stations for a given month and time are read at once. Decoding of 32-bit words is done individually.

c. Perform objective analysis and bilinear interpolation if required. If the requested location matches a station in the station data file, then no objective analysis or interpolation is required. If there is no match, then the analysis will be performed to estimate the unconditional probability at that location.

d. Compute a joint probability between two unconditional probabilities if required. If the user requests a joint probability, WWMX first estimates the unconditional probabilities as in paragraph c above for the requested location. Then the joint probability is estimated using an empirical function developed by Boehm (1974):

$$\text{Joint Probability} = 0.7\ (P_C P_V) + 0.3\ \text{minimum}\ (P_C \text{ or } P_V)$$

where P_C = probability of ceiling greater than threshold
P_V = probability of visibility greater than threshold
Joint probability = probability of P_C and P_V greater than thresholds

e. Write results. Output is to remote terminal or disk-file in a format suitable for dump to a printer, in self-explanatory form. There is a specialized output capability for the purpose of generating an input file compatable with USAFETAC program ADXOSCN. ADXOSCN has the ability to geographically display the results of the selected analysis routine and the station data. (See Appendix A).

CHAPTER 5

OBJECTIVE ANALYSIS

5.1 Introduction

Objective analysis deals basically with the problem of interpolating data from a set of irregularly distributed reporting points (stations) in order to assign estimates of a variable (ceiling, visibility, etc.) to a regular grid network. An objective analysis scheme performs several functions. A first guess field for the grid point values is found first. Then successive corrections at the grid point are done and finally the field is smoothed (Cressman, 1959).

The following sections contain the three types of analysis algorithm that were investigated. They are the Barnes analysis, the Janota analysis, and a nearest neighbor analysis. Each one of these analyses uses all or part of the objective analysis system proposed by Cressman. The theory and usage of each of these analysis schemes will be discussed and ways of optimizing each will be proposed.

5.2 Barnes Analysis

5.2.1 Theory. The Barnes technique is designed to analyze accurately small variations in a data field without excessively amplifying the noise inherent in it (Janota 1966). To accomplish this, Barnes makes the fundamental assumption that a two-dimensional distribution of an atmospheric variable can be represented by the summation of an infinite number of independent harmonic waves, a Fourier integral representation.

Using the assumption that an atmospheric quantity can be represented by a Fourier integral, one may define a corresponding smoothed function, which is obtained by applying a filter to the original function.

$$g(x,y) = \int_0^{2\pi} \int_0^{\infty} f(x + r\cos\theta,\ y + r\sin\theta)\ w r dr d\theta \qquad (4)$$

w is defined as the weight factor or filter and is

$$w = (1/4\pi k)\ \exp\ (r^2/4k) \qquad (5)$$

r and θ are polar coordinates with the origin being at the point (x,y), and k is a parameter determining the shape of the weight factor.

One may rearrange (4) to express the weight factor in an alternate form. This is done since in (4) the maximum weight is not applied to r = o.

$$g(x,y) = \int_0^{2\pi} \int_0^{\infty} f(x + r\cos\theta,\ y + r\sin\theta)\ x(\eta/2\pi)\ d(r^2/4k)\ d\theta \qquad (6)$$

$$\eta = \exp\ (-r^2/4k) \qquad (7)$$

Rearranging Equation (4) solves the problem of applying the maximum weight at r = o, but another problem has arisen. Interpolation by Equation (6) is not practical because first, the analytical form of f(x,y) is not known. In fact, that is what must be represented by a few random pieces of information. Second, the function cannot be integrated to infinity. It must be approximated by placing a finite limit on the region of influence of any piece of data. Also, one must take a weighted average of data within that region of influence.

$$g(x,y) = \frac{\sum_i^N \eta(r_i) f_i}{\sum_i^N \eta(r_i)} \qquad (8)$$

The above Equation (8) is the practical form of Equation (6). The technique using the above equations is fully described by Barnes (1964), but is simply stated by Janota (1966) as follows

a. Obtain a weighted average of the data using a given radius of influence and the weighting function. Perform an initial interpolation over the grid. This is the first guess field.

b. From this grid-point analysis, interpolate a value ϕ_a at each data point and compute the error of the analysis. ϕ_d is the observed value.

$$\phi_d - \phi_a = e \qquad (9)$$

c. Analyze the error field using the same radius of influence and weighting function.

d. Add the computed error field to the first guess field to obtain a new, more detailed analysis.

e. Continue iterating steps b through d until the residual has been diminished to the desired amount of detail.

This method effectively dampens the growth of shortwave or noise components while permitting synoptic scale features to be represented adequately. No additional filtering is required to achieve this result.

5.2.2 Usage. The Barnes analysis subroutine used in this project was originally written for USAFETAC by Major Arnold Friend in support of Reforger '78. It was modified slightly for this project.

The subroutine that executes the analysis is supplied with the GWC 1/2-mesh super-grid coordinates of all the stations in the data set, the probability of the requested weather parameter for a certain month, time, and threshold for those stations, and the actual size of the grid being used. It returns a two-dimensional array, which is dimensioned to the number of grid units in the I-direction by the number of grid units in the J-direction. This array contains the probabilities of the requested weather parameter at each grid point.

At the beginning of the subroutine, two variables need to be initialized: first, the number of iterations that the analysis does; and second, the radius of influence a given point has on its surroundings. Next, it takes each report and if the data is not missing, it evaluates the best estimate within the given radius of influence at each of the grid points. This is accomplished by truncating the stations' locations to the upper left corner of the grid block. Then the distance from the upper left corner of each grid block to the station is computed. Next, a bilinear interpolation is done to find values at the station based on the current estimate at each of the four surrounding corner grid points. After this is done, determine which grid points are within the radius of influence of the data. At this time, the appropriate corrections and weights are applied to each grid point. First, the distance from the data point to the grid point is computed. Next, eliminate computations performed at the corners where the weight factor is negligible. Finally, the sum of all the corrections and weight factors applied at every grid point within the radius of influence of the grid point being considered is computed. The best estimate at each grid point is computed after the scan. This whole process is repeated depending on the number of iterations requested.

5.3 Janota Analysis

5.3.1 Background. Paul Janota asked the following questions in AWS-TR 188, "What type of information does the customer require at an analysis grid point? What are the characteristics of the data? What scales are inherent in the variable being analyzed? What is the grid scale required for the final depiction?" From these questions he compared various analysis methods and reported his conclusions about methods of objective analysis. From these conclusions, Lt Col Peter Havanac, USAFETAC/DN, wrote a computer routine in support of Project 2304 (Electro-Optical Data Base) to do an objective analysis on a discontinuous vari-

able field. That routine has been substantially modified for USAFETAC/DND Project 2502.

5.3.2 Usage. As stated earlier, this program does an objective analysis on a discontinuous variable field. As in the Barnes analysis, this analysis is supplied with the AFGWC super-grid coordinates of all the stations in the data set and the probability of the requested weather parameters for a certain month, time, and threshold for those stations. It also returns a two-dimensional array containing the probability of the requested weather parameter at each grid point. From this, one can determine the probability of the weather parameter requested for the requested location.

First, an initial sort is done to see if any stations in the data set lie within a given distance of a grid point. If no stations are found, the search radius is increased by the square root of two. If only one station is found, its value is assigned to that grid point. If just two stations are found, then the value of the closest station to the grid point is assigned to that grid point. If three or more stations are found then the data is checked for bimodality. If the data is bimodal, it assigns the grid point the value of the closest station. If the data is not bimodal, it assigns a distance weighted average of the station values to the grid point. This process is repeated until the two-dimensional array has been filled. At this point, the array is passed back to the control subroutine and an interpolation is done on this array. The interpolation finds the probability of the given weather element for the requested location.

5.4 Nearest Neighbor Analysis

5.4.1 Background. This routine was created to supply a quick method for the determination of the probability of a requested location for a certain month, time, and threshold for a given weather parameter. This is accomplished by taking the square root of the sum of the squares.

5.4.2 Usage. Initially, the program is given as data the position, in AFGWC super-grid coordinates, and the probabilities of the requested weather parameter for all the data stations. It is also given the latitude and longitude of the requested location.

With this information the program then calls a subroutine that "degrids" the latitude and longitude of the requested location into AFGWC super-grid coordinates. Next, the degridded location is compared with all the data stations. If a data station is within one grid unit of the requested location, it is held as a possible neighbor. If no stations are found with the original search radius, then the search radius is increased by one grid unit until at least one station is found.

The actual neighbor is found by the Pythagorean theorem, squaring the I-value, adding it to the squared J-value, then taking the square root of that value. The resulting number is the distance of that station from the requested location. Finally, the program finds the smallest distance from the requested location. It then finds the probability of the station that is closest to the requested location, and assigns this value to the requested location as the probability of that weather parameter at that location for that month, time, and threshold.

5.5 Discussion

The three analysis schemes used in this project were either adapted from previously written software or the nearest neighbor routine written specifically for this project, e.g., each one of the analysis schemes can be optimized to produce better results. This section discusses the ways the analysis schemes were optimized.

The Barnes analysis in theory is not limited to any particular distribution of data, but as Barnes (1964) points out applications should be made to reasonably uniform data distributions. This is for economical reasons because a smaller radius of influence can be used and thus the scheme converges more quickly.

In optimizing this scheme, as was the rule for the other schemes, the relative central processing unit (CPU) time was weighed against the improvement of the root-mean-square error (RMSE). Three major areas in which the Barnes analysis can be optimized are:

a. Vary the number of iterations the analysis accomplishes.

b. Vary the radius of influence.

c. Add "pseudo" data to fill data sparse areas.

Our effort was restricted to the first two options. Many questions remain unanswered at this time concerning the feasibility and effectiveness of using pseudodata to improve results.

In the optimization, several runs of the program were done. The radius of influence was varied from three to five grid units and the number of iterations was varied from three to seven times. The procedure was as follows:

a. Set the radius of influence and number of iterations.

b. Run the program and gather the needed statistics.

c. Hold the radius of influence constant and vary the number of iterations.

d. Run the program again and gather the new statistics.

e. Do the above for all the specified iterations.

f. Change the radius of influence and repeat steps b through e.

It was found that 3.5 grid units as a radius of influence and 4 as the number of iterations was the best combination for this data set.

Figure 6 is a sample plot of the experimental area using the Barnes analysis. Initially, it was found that there was a significant boundary value problem. Thus, a plot of the field would often have spuriously analyzed fields at the boundaries. To reduce the impact of this problem the grid values were initialized to the mean probability of the requested weather parameter. This also caused the program to converge more quickly.

The Janota analysis was considered for this project because it includes a check for bimodal data and it effectively preserves discontinuities in the field. If the data is bimodal in the influence region the closest station to that point is used to supply the probability. As Janota (1966) pointed out, the total cloud cover has a characteristic bimodal frequency distribution. He also stated that the discontinuity-preserver would better describe the shape and intensity of cloud cover and eliminate the artificial intermediate cloud amounts.

Figure 7 is a sample plot of the experimental area using the Janota analysis. This plot uses the same month, time, and threshold as was used for the plot of the Barnes analysis.

The nearest neighbor routine was created to supply a quick method for the determination of the probability of a certain weather parameter for a requested location. Using this method the RMSE was about 14 percent. One way of reducing the error would be to add a 'smoother.' Using a smoother as described by Fleming (1969) one can reduce the RMSE to below 8 percent. This alone would bring the RMSE closer to that of the Barnes analysis (5-6 percent) and the Janota analysis (approximately 7 percent). Employing the smoother will increase the CPU time, but the reduction in RMSE would be substantial enough to warrant its usage.

The advantage of a strict nearest neighbor routine is that it has a tendency to preserve discontinuity. If the smoother is used one loses some of the discontinuity in return for a smaller RMSE.

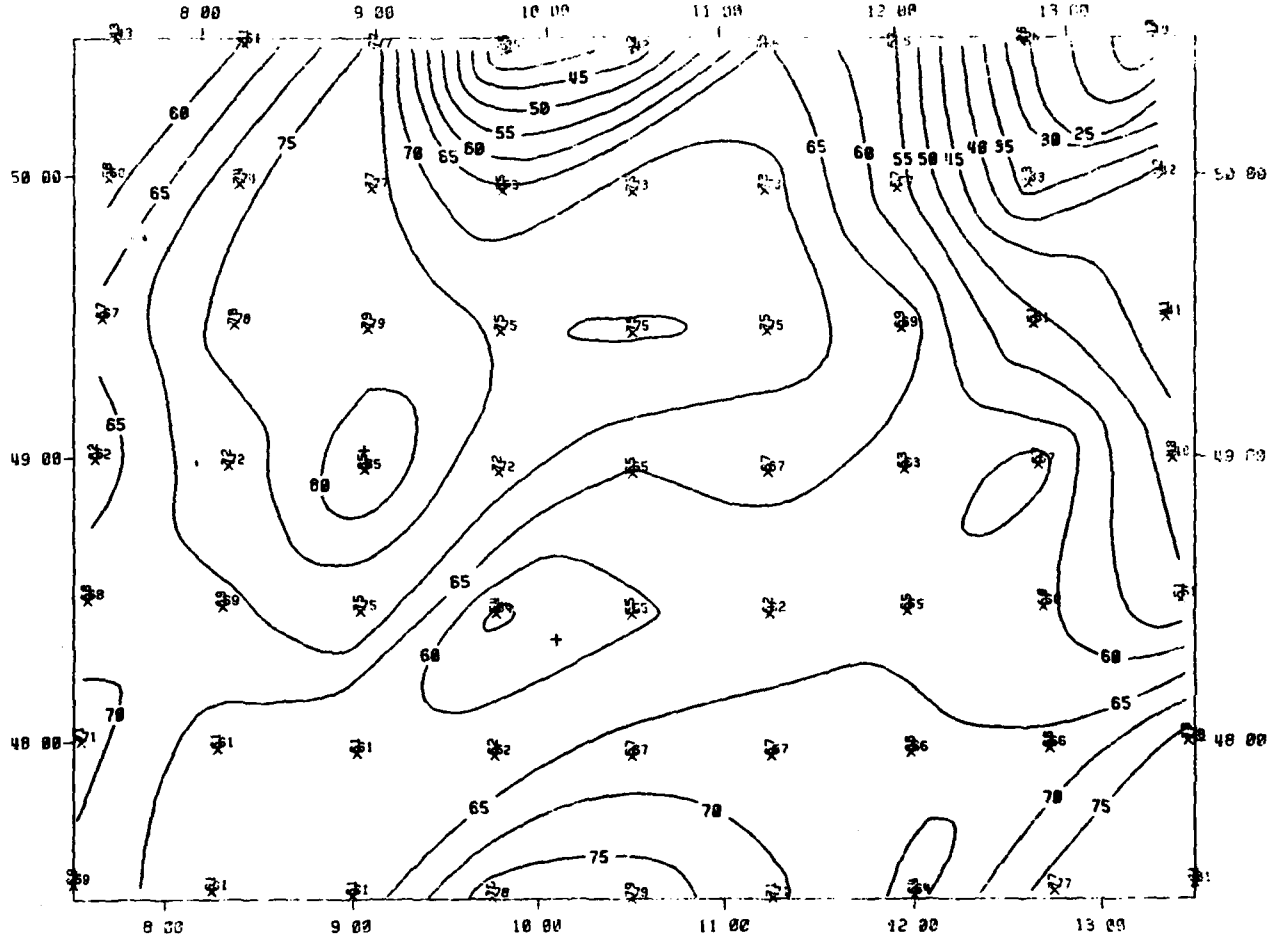


Figure 6. Barnes Analysis, October, 05-07Z, Ceiling .GT. 1000FT AGL.

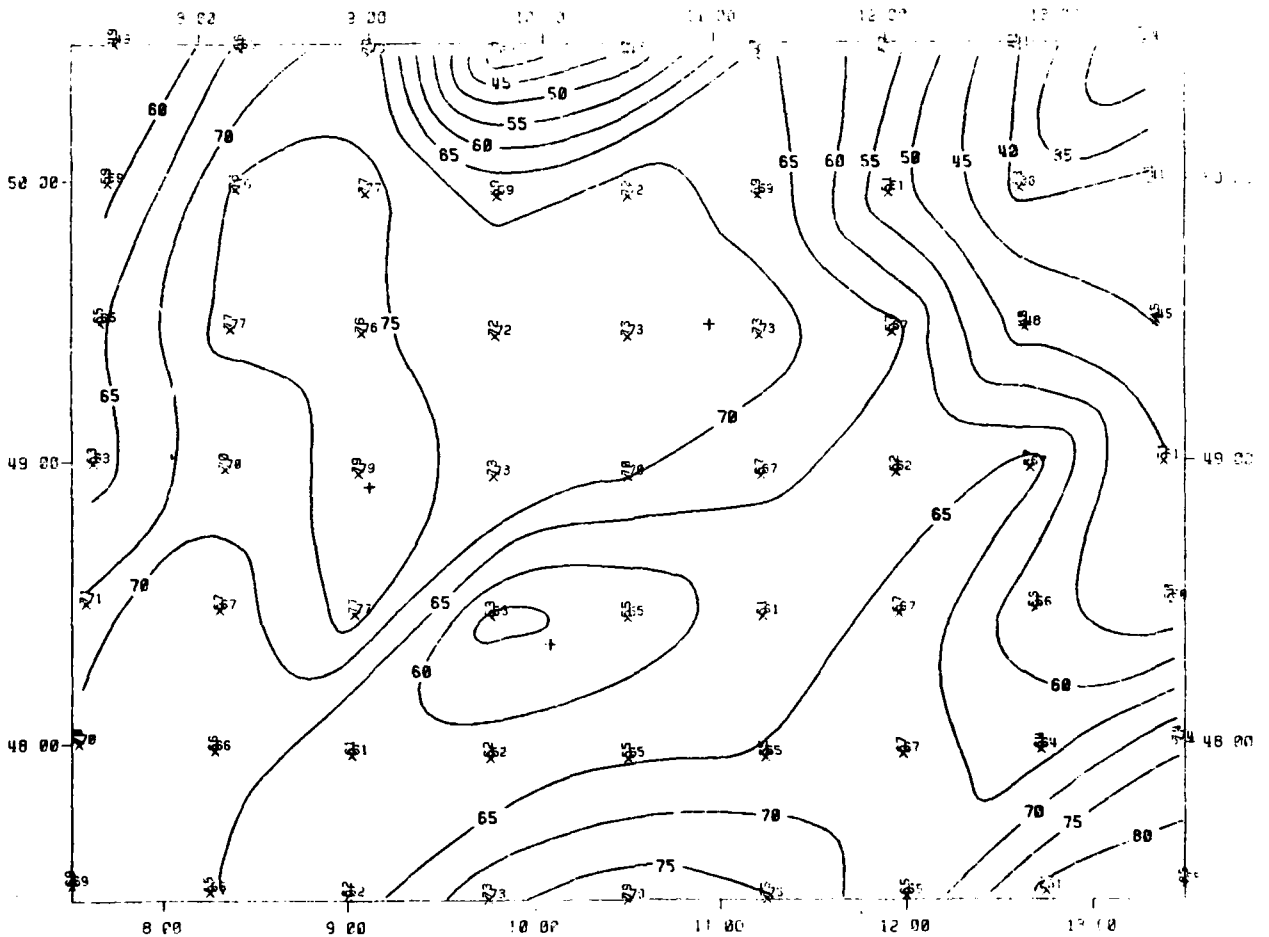


Figure 7. Janota Analysis, October, 05-07Z, Ceiling .GT. 1000FT AGL.

Figures 8 and 9 are sample plots of the experimental area using the strict nearest neighbor routine and the nearest neighbor routine using the smoother.

It was found that the Barnes analysis was the superior objective analysis routine. This and other findings will be discussed further in Chapters 6 and 7, the results and recommendations chapters, respectively.

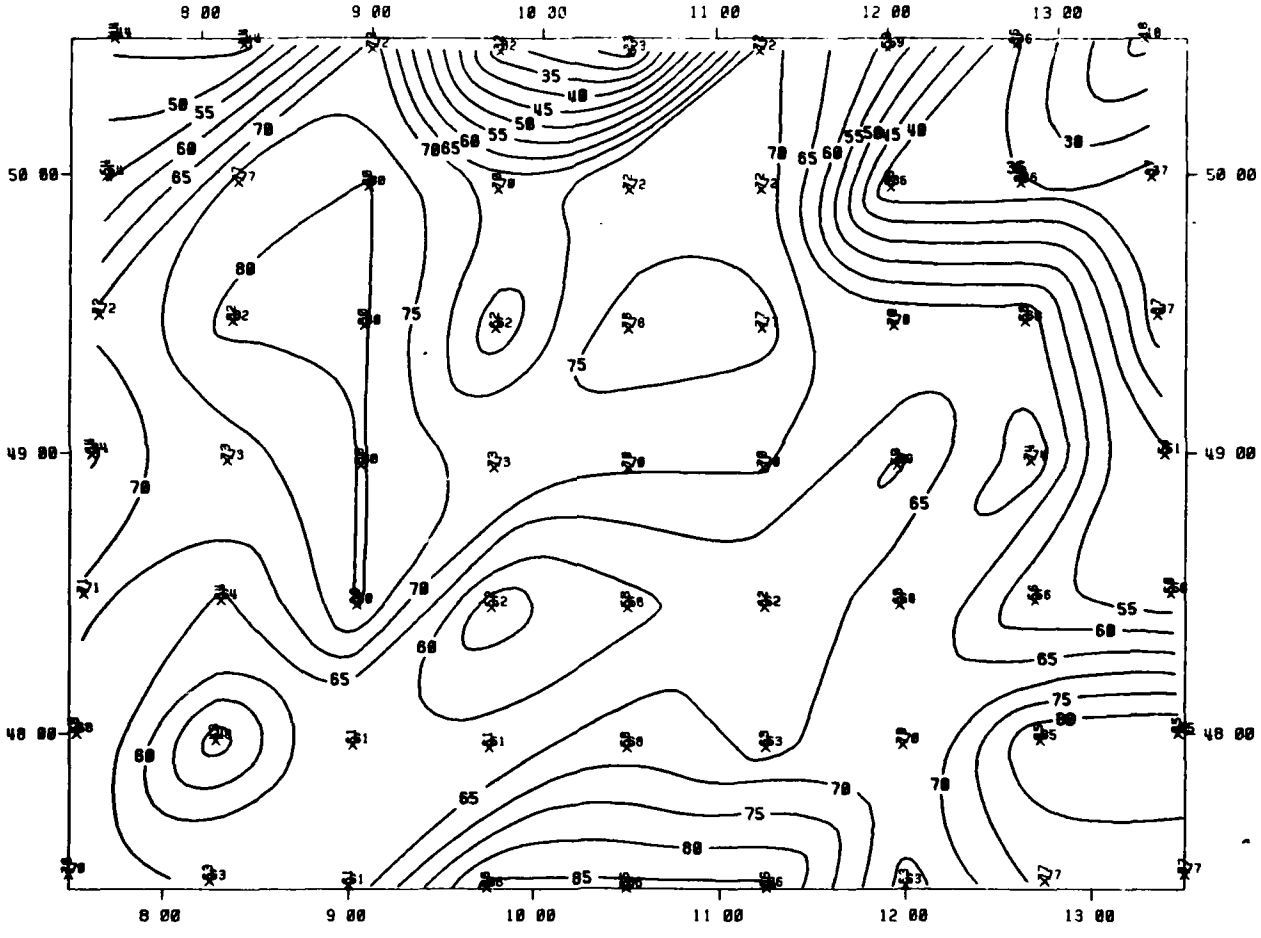


Figure 8. Nearest Neighbor, October, 05-07Z, Ceiling .GT. 1000FT AGL.

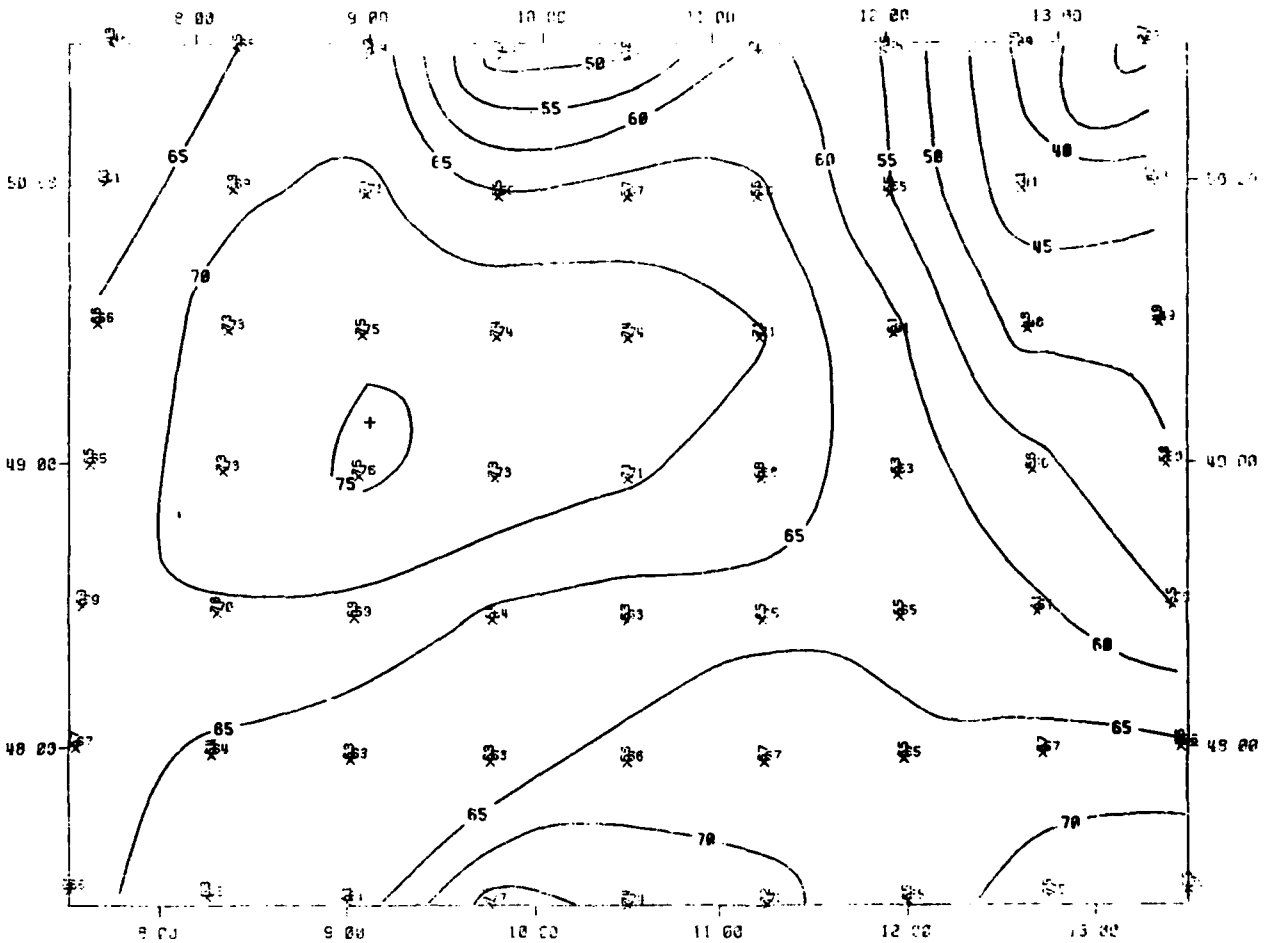


Figure 9. Nearest Neighbor Using the Smoother, October, 05-07Z, Ceiling .GT. 1000FT AGL.

CHAPTER 6

RESULTS

6.1 Fit of Unconditional Ogive

The quality of a line-segment-selection fit depends on two factors. First, the number of line-segments used has an effect. The more segments that are used, the better the fit. However, the use of more segments requires more storage space. Secondly, the "standard" thresholds chosen have an impact on the quality of fit. The single set of standard thresholds that were chosen worked well for most of the 81 stations in the data file. The data from some stations did not fit as well when using the standard thresholds described in Chapter 2. For this reason, the ability to store and retrieve standard categories specific to a station was incorporated into this technique. However, no attempt was made to optimize the standard catagories for individual stations for this project. This would, of course, be included in any continued future effort.

Without any such attempt to optimize the fits, the results were still outstanding. Of 71 stations used to test the technique (approximately 13,000 fits) only five had an overall RMS difference in excess of 1 percent. The statistics of goodness of fit are summarized in Table 1.

Table 1. Line Segment Selection -- Quality of Fits.

	RMS Difference (Percent)	Avg Max Diff Per Fit (Percent)	Standard Deviation Max Diff (Percent)	Sample Size
Ceiling	0.71	2.3	0.63	~6500
Visibility	0.54	1.4	0.42	~6500

The accuracy of the fits was so convincing that no need was found to compare the results of the objective analysis techniques to the original data in uncompacted form; the compacted unconditional probabilities differed in no significant way from the originals.

6.2 Modeling Joint Probabilities

The computation of joint probabilities from the unconditional probabilities saves more space then any other facet of the entire data compaction scheme. For this reason some discussion on the errors incurred by the use of the technique is in order. No rigorous statistical test was made to verify the quality of the technique. The technique is accepted and used by analysts in USAFETAC Aerospace

Sciences Branch to compute joint ceiling/visibility probabilities from unconditional probabilities. An experiment was performed using Frankfurt data as an example. The Frankfurt experiment consisted of verifying the WWMX CIG/VIS model against a published RUSSWO for Frankfurt. The RUSSWO itself had been computed using a completely different period of record. Two-hundred joint probabilities were generated for five arbitrary threshold pairs at twice a day (02-04Z and 14-16Z) for four months. Eighty unconditional probabilities were also generated. No objective analysis was performed. The results are summarized in Table 2.

Table 2. Joint Probability Verification -- Frankfurt Experiment.

	RMS Difference (Percent)	Avg Max Diff Per Month/Time (Percent)	Sample Size
Unconditional (Edge)	2.02	4.23	80
Joint (Interior)	3.16	6.58	200

This test and others performed by these analysts suggest that errors of computed joint probabilities are somewhat greater than the errors in the estimated unconditional probabilities. These tests indicate that the empirical function, when used to compute the joint probability from unconditional probabilities with an assumed zero error, reproduce the interior joint probability with an RMS difference in the vicinity of 1 percent. It is estimated the error introduced by the computation of joint probabilities ranges from 1-3 percent RMS difference, depending on the station.

6.3 Selection of Objective Analysis Method

In Chapter 5, the three objective analysis techniques which we investigated for this project were discussed. These were the Barnes, Janota, and nearest neighbor. The performance of each of these techniques was thoroughly evaluated. The Barnes algorithm clearly worked best for the purpose of this project.

The objective analysis quality evaluation was itself performed objectively and on independent data. For each "synoptic" month and time group, ceiling and visibility ogives were available for each of the 81 stations (although some stations might be missing for a particular month and time). For each such "map" time, the data field was objectively analyzed purposely omitting one station from the analysis. Then a comparison of the an estimate of the probability for that missing station location with the actual value was performed. The difference between the estimate and the actual station value, i.e., the residual, was retained and the process repeated for each of the 81 stations for each "map" time.

All 12 months and all 8 times of day were evaluated. The ceiling thresholds that were used were 200, 1000, 2000, 3000, and 10,000 feet. The visibility thresholds were 0.5, 1, 2, 3, and 5 miles. It was possible to stratify these statistics by month, time, threshold, station, and by arbitrary groups of stations. Over 200,000 analyses of the data field were performed to obtain these statistics.

It was determined that the Barnes algorithm worked best for these parameters. The Janota method was slightly worse, and the nearest neighbor method was by far the worst. The overall analysis results of these methods is shown in Table 3.

Table 3. Comparison of Objective Analysis Methods. Overall analysis results--all months, times, and thresholds.

		BARNES	JANOTA	NEIGHBOR
Ceiling	RMS Difference	5.7	7.6	15.2
	Max Difference	44.1	78.4	81.2
Visibility	RMS Difference	5.7	7.5	13.7
	Max Difference	41.6	66.2	68.8

It became clear that these climatological probability data were to a great extent conservative and fairly continuous. This was why the Barnes method worked best. The Barnes analysis tended to smooth the field more than the other two. This seemed beneficial and supported our decision to use an interpolation model rather than a regression model.

6.4 Performance of the Barnes Method

Looking more closely at the Barnes performance in Figure 10, one can see the method varied month by month. Ceiling statistics were best in winter and worst in late spring and early summer. The reason for this appears to be that cloud patterns are much more systemic during the winter season causing the climatological probability field to be more uniform and continuous. The visibility statistics were best in summer and early autumn. Visibility tends to be good everywhere during that time of year--again causing the field to be more continuous.

Figure 11 shows the variation of accuracy with time of day. Both ceiling and visibility were most accurate during the middle of the night (0100-0300 LST). Visibility actually varied little with time of day, while ceiling displayed a definite minimum of accuracy in midafternoon. This is most likely a reflection of localized convective activity, particularly during summer.

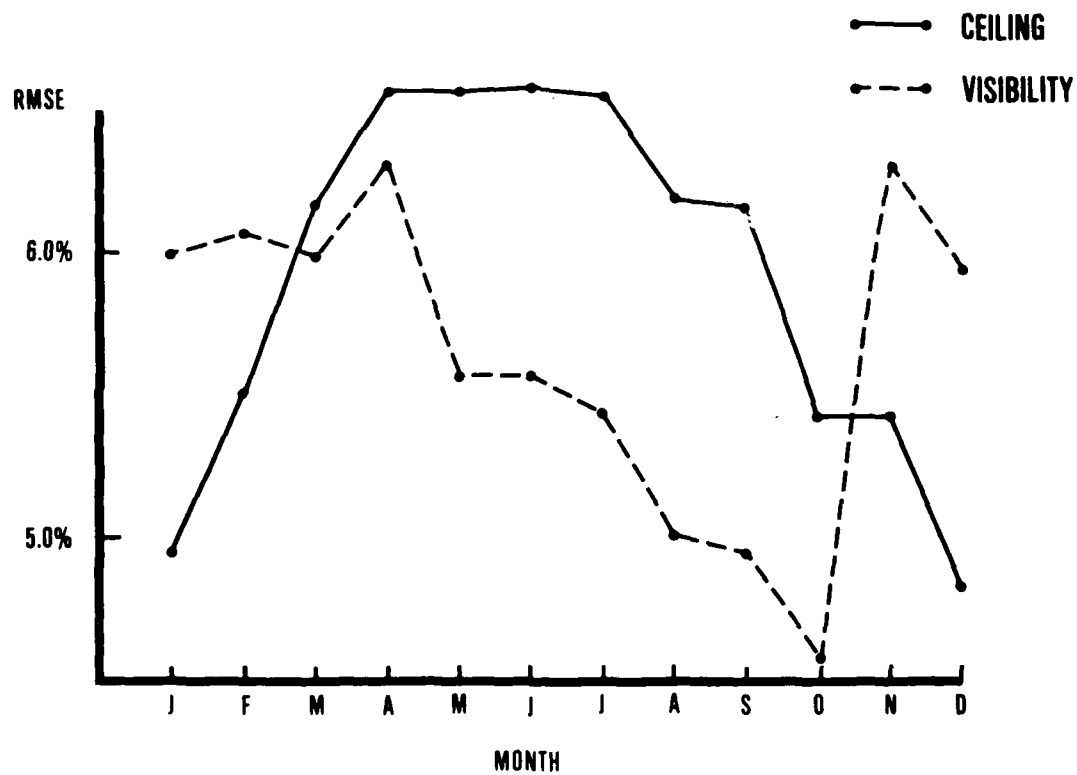


Figure 10. Variation by Month.

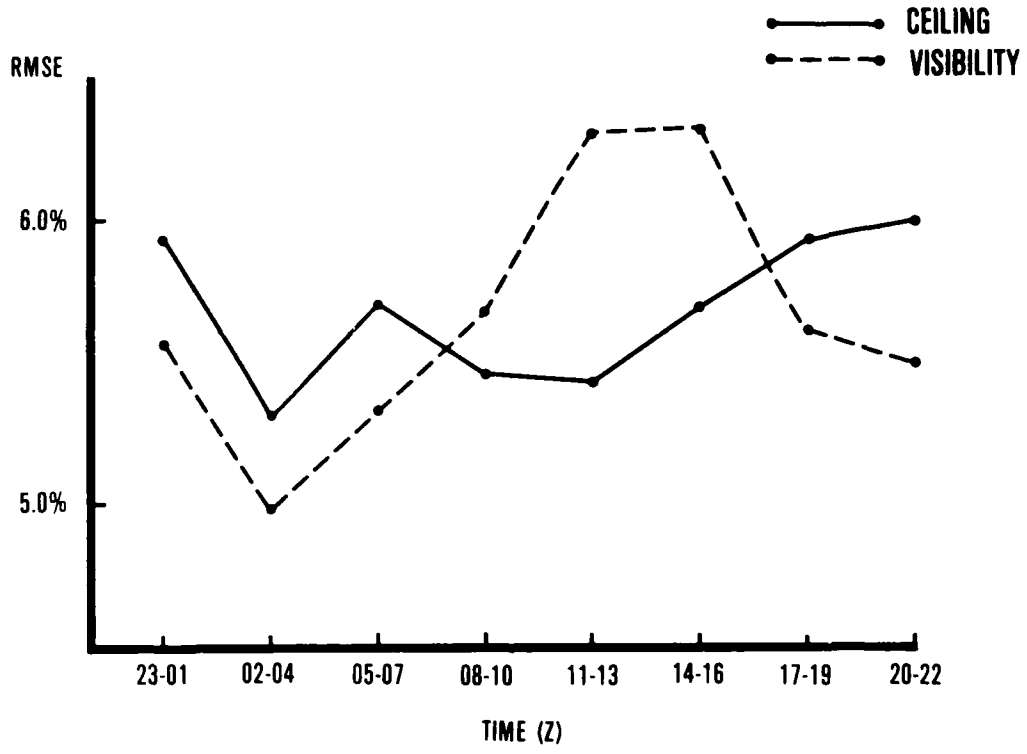


Figure 11. Variation by Time of Day.

There was very little pattern observable as variation by ceiling or visibility threshold. The method seemed to work slightly better at very low thresholds and then again slightly better at higher thresholds (above 3000/3).

Of course, the greatest variation was geographical in nature. Figures 12-16 display the performance of the method over the area studied in this project. Figure 12 shows the typical error for any given map time. Figures 13 and 14 display the RMS difference for ceiling and visibility for all months, times and thresholds. Figures 15 and 16 show the absolute maximum error for all realizations--over 400 per station.

In each map a typical pattern is observed. The best performance is in the central plains regions. Only slightly worse is the performance in the northern and southern mountain stations (higher than 500 meters). The greatest errors tend to be grouped consistently in the boundary regions of the experimental area. The actual statistics for these three groups of stations are shown in Table 4.

Table 4. Variation by Selected Station Groups.

	RMS Difference (Percent)		
	CENTRAL PLAINS (41 STNS)	MOUNTAINS (20 STNS)	BOUNDARY (20 STNS)
Ceiling	4.0	5.7	8.1
Visibility	4.2	5.5	8.0

The apparent boundary problem was primarily caused by three nontypical stations which caused large errors for their neighbors. These stations happened to be in the boundary regions of our area.

6.5 Computation of Joint Probabilities at Nonstation Locations

The effectiveness of the technique in computing joint probabilities at a location for which we actually had no data was evaluated. We were able to obtain a RUSSWO for Hohenfels, a station for which we had no data in our file. The interior probabilities were again computed for 200 ceiling/visibility pairs for Hohenfel's location and then were compared to the RUSSWO probabilities. There was a 7.5 percent RMS difference and an average maximum difference per RUSSWO page of 15 percent. Recognizing that one station is too small a sample from which to generalize, it appears that we can evaluate the joint probabilities at nonstation locations about as well as the unconditional probabilities. The RUSSWO itself was of highly variable quality. For some month-times there were fewer than 100 observations, which for others more than 700. This technique worked significantly better for the times which had high observation counts.

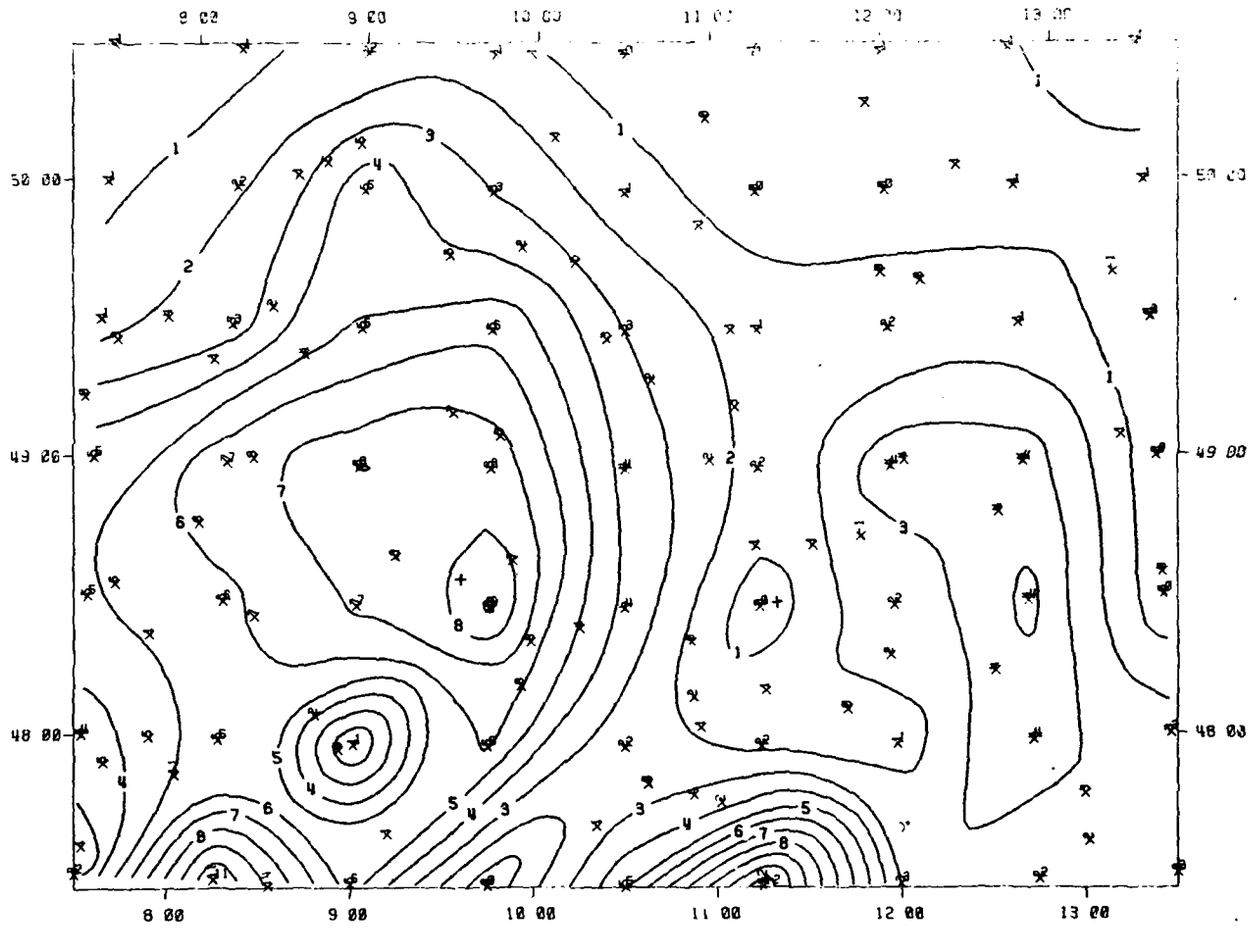


Figure 12. Estimated Residual for Ceiling .GT. 3000FT, January, 17-19Z, Southern Germany.

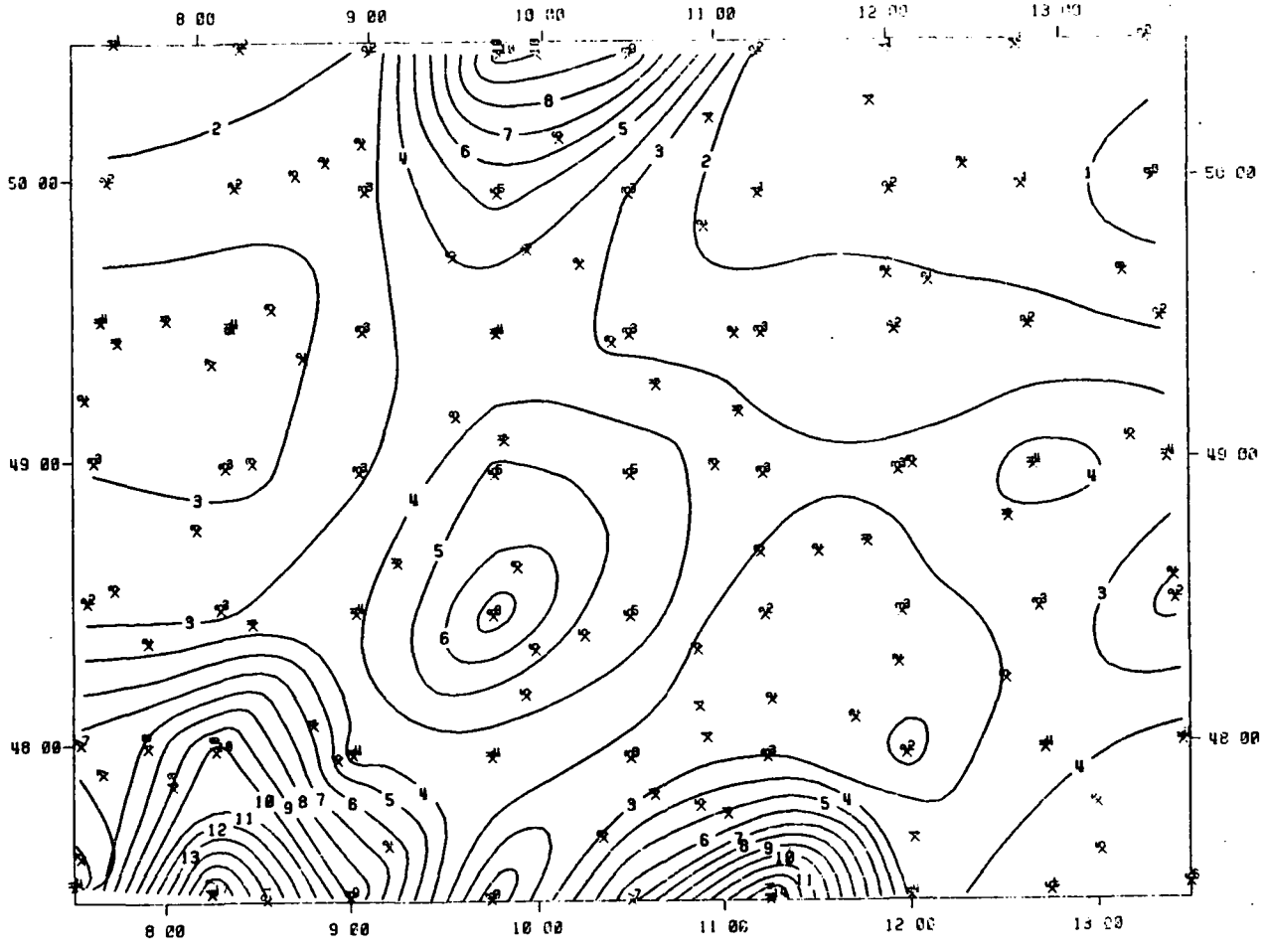


Figure 13. RMSE Difference Field, Ceiling, Annual, Southern Germany.

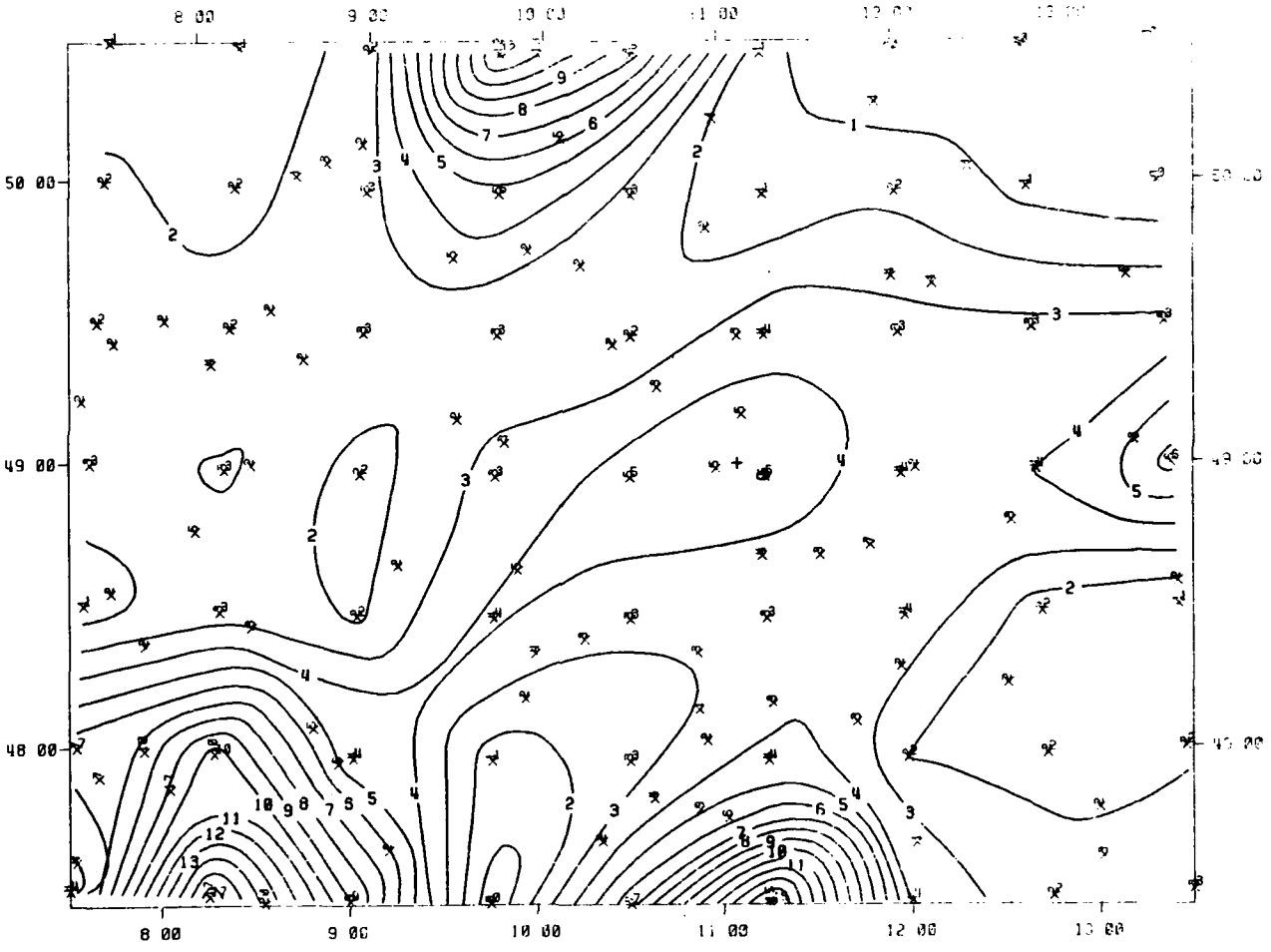


Figure 14. Estimated RMSE Difference Field, Visibility, Annual, Southern Germany.

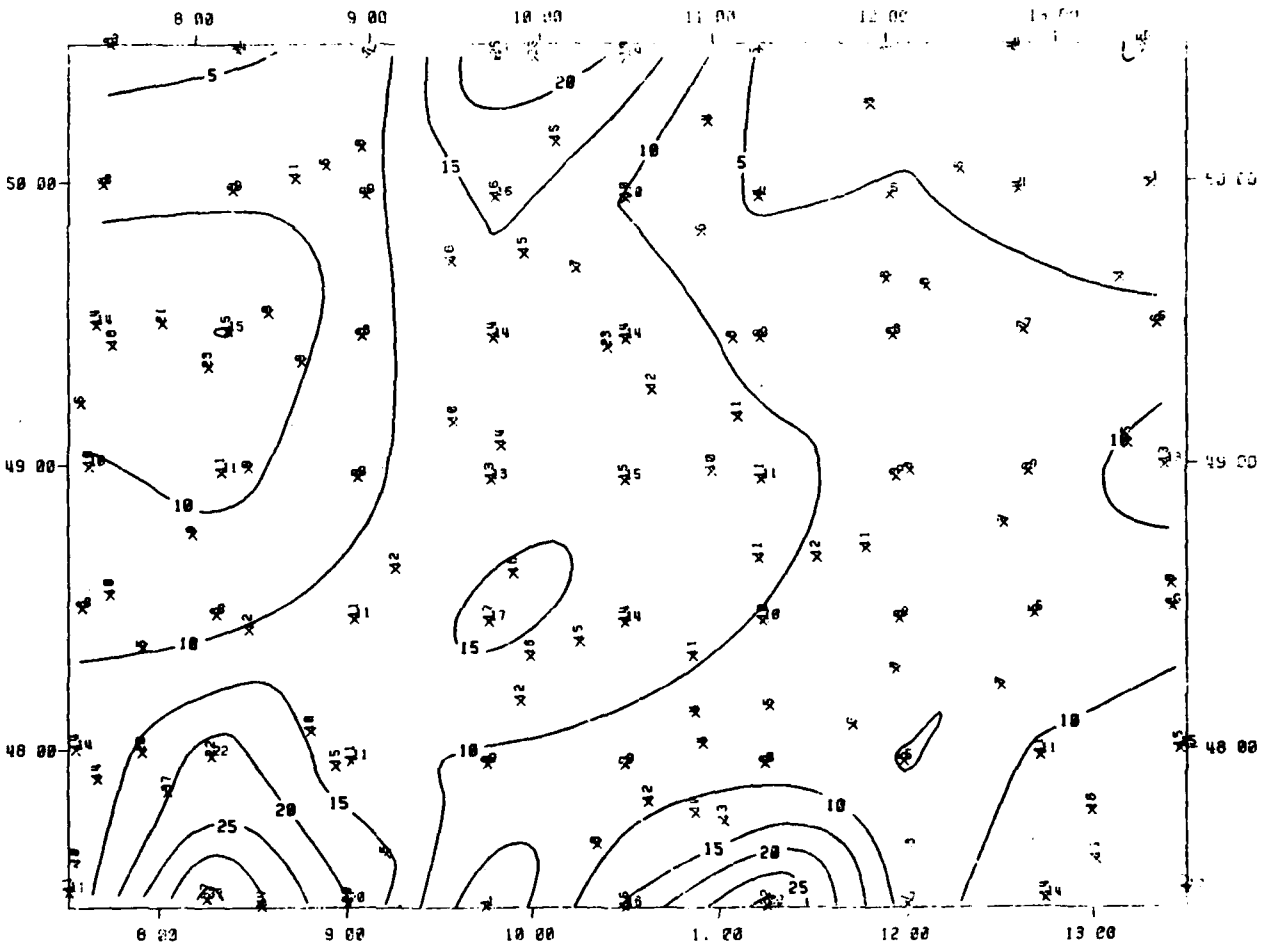


Figure 15. Estimated Maximum Error Field for Ceiling, Annual, Southern Germany. (Based on all months, times, and selected thresholds.)

Figure 16. Estimated Maximum Error Field for Visibility, Annual, Southern Germany. (Based on all months, times, and thresholds.)

Chapter 7

CONCLUSION

7.1 Factors to Consider

In arriving at our conclusions concerning the future course for USAFETAC in the area of climate data interpolation, we considered many factors. First, has a capability which promises to be operationally useful actually been demonstrated? We say yes. Although the results of the interpolated data are certainly less than perfect, they are better than might have been expected considering the innumerable complicating factors. These results are superior to any previous effort that we are aware of. It must also be emphasized that the "errors" mentioned in this discussion are artificial in the sense that they were errors in information that is not actually available--information which is in a sense nonexistent and unobtainable until now. Thus, the real question: is there sufficient information contained in these estimates to make them operationally useful? Unfortunately the answer to that question is outside the realm of the knowledge or experience of these analysts, although it has been said that "answers" within 10 percent would be "good enough." This is a question which will have to be directed to the customers.

There is another consideration: is this the best type of model to pursue? Again our response is yes. The interpolation model is simple and anchored to the actual data. Customers can better understand how their information was obtained. Furthermore, at least these preliminary results indicate that it works better than other possible model types. A related question is concerned with how this model would work in data-sparse and data-void regions. Here the answer must be - very poorly. This problem has not been solved. However, the idea of "pseudo-stations" or "pseudodata" is proposed. These would be fictitious or "made-up" data points, derived by some means and stored just as actual station data is stored now. The study of the feasibility of this idea would be a major part of the initial work in the continuation of this project. Perhaps some type of geoclim regression model would make a first guess at the distribution which could then be hand-massaged subjectively by a "wise and learned" analyst. In any case, this is a question requiring further careful study.

A less technical question, but a very important one, is how much manpower can be devoted to this effort, and for how long? Here it can only pointed out that there are two USAFETAC slots dedicated to WWMCCS, so the question really becomes, is this how these slots should be employed?

Finally, what will be the USAFETAC computer configuration in the future? Fully implemented, this technique will require a fair amount of on-line disk,

approximately 40 megabytes. In addition, to be most operationally effective, an interactive capability would be necessary. These needs would be effectively met by the procurement of the hardware requested in the Enhancement DAR.

7.2 Actions

With these factors in mind USAFETAC/DND will take the following actions:

a. Begin immediately to implement this technique for ceiling and visibility for the European Theater. The primary cost of this implementation will be conversion of the software to the IBM 4341. These costs have been minimized by keeping the IBM in mind throughout the technique development. Much of the conversion cost will be in software documentation which has been kept to a minimum in the volatile technique development phase of the effort. The bulk of the initial data collection for this implementation has already been accomplished by DNO for the AFGWC MOS Project 1564. It is estimated that this implementation will take approximately one man-year. A decision to add additional areas will be made prior to the conclusion of this effort.

b. Begin to investigate the feasibility of the pseudodata concept. Results of this investigation will be crucial to the future decision whether to implement these techniques on a worldwide scale. Any attempt to produce such pseudodata should be semiobjective in nature and, perhaps, should incorporate circulation-regression techniques such as the Geoclim model used in support of Reforger '76. This preliminary investigation will also cost approximately one man-year.

A capability to provide objective climate probability estimates for anywhere in the world is a subject which has been discussed at USAFETAC and in the field for over a decade. There is a real demand and need for such a capability. There is yet a long way to go. However, these analysts believe a feasible way to attack the problem has been demonstrated by this project. The development and operational implementation of these techniques will very soon begin to meet some of the stated requirements of WWMCCS customers as well as many others.

REFERENCES AND BIBLIOGRAPHY

Barnes, S. L., 1964: "A Technique for Maximizing Details in Numerical Weather Map Analysis," J. Appl. Meteorol., Vol. 3, No. 8, pp. 396-409.

Bean, S. J., M. Hensen, and P. N. Somerville, 1981: "A Fortran Program for Estimating Parameters in a Cumulative Distribution Function," AFGL-TR-81-0120, 31 March 1981, 15p (AD-A104165).

Boehm, Albert, 1974: "AWS Technique Development Summary," Hq Air Weather Service, Scott AFB, Illinois, September-November 1974, pp. 16-17.

Boehm, Albert, 1977: "Optimal Decisions Through Mission Success Indicators," in: Proceedings of the 7th Technical Exchange Conference, 30 November-3 December 1976, published by Atmospheric Sciences Laboratory, White Sands Missile Range, New Mexico, April 1977, pp. 17-25.

Boehm, Albert, 1979: "The Rank Input Method and Probability Variation Guides," USAFETAC/TN-80/004, July 1980, 13p (AD-A093196).

Cressman, G. P., 1959: "An Operational Objective Analysis System," Monthly Weather Rev., Vol. 87, No. 10, pp. 367-374.

Fleming, R. J., 1969: "Air Force Global Weather Central Fine Mesh Upper Air Analysis Model," AFGWC TM 69-2, 15 December 1969, 19p (AD-710203).

Friend, A. L., 1978: "An Objective Technique for Spreading Climatology," USAFETAC-PR-78-007, USAF Environmental Technical Applications Center, Scott AFB, Illinois, June 1978, 4p.

Glahn, H. R., 1981: Comments on "A Comparison of Interpolation Methods for Sparse Data: Application to Wind and Concentration Fields," J. Appl. Meteorol., Vol. 20, No. 1, January 1981, pp. 88-91.

Janota P., 1966: "Some Objective Analysis Techniques Suitable for Nephanalysis," in: "Technique Development Reports - 1966," AWS TR 188, November 1966, pp. 1-16 (AD-654118).

Martin, D. E.and E. Myers, 1978: Climate Models That Will Provide Timely Mission Success Indicators for Planning and Supporting Weather Sensitive Operations," AFGL-TR-78-0308, December 1978, 562p (AD-A066796).

APPENDIX A

MAPS OF ESTIMATED PROBABILITY FIELDS

A.1 Considerations in the Interpretation of Maps

Maps in this appendix were created using USAFETAC program ADXOSCN. ADXOSCN contours gridded latitude/longitude data. It does not perform any objective analysis of irregularly spaced (station) data. To obtain a set of gridded latitude/longitude data, program WWMX is requested to provide estimates of probabilities at the latitude/longitude grid points. Thus, the contour lines on the maps apply to the estimated probability fields, not station data. All the maps in this appendix show probability fields estimated using the Barnes algorithm. The points in the estimated probability field are denoted by an × with the estimated probability (to the nearest percent) plotted both vertically and horizontally $\overset{91}{\times}$. It is this set of points that ADXOSCN contours. One should realize that there are any number of ways to objectively draw contour lines through such a field of grid points. ADXOSCN provides the user with some flexibility in the specification of the contour characteristics. These maps were all produced with the same contour specifications, which result in relatively smooth contours.

To further enhance an analyst's ability to interpret the estimated unconditional probability field, the actual probabilities at stations are superimposed on the contour field. A station value is denoted by an × with the probability plotted only vertically $\overset{86}{\times}$. A $\overset{--}{\times}$ indicates that data for that month/time at that station is missing. The total number of stations in the map area and the number of those missing data are indicated at the bottom of the map. The number of missing "observations" subtracted from the total number of stations in the area, is the number of data points used to generate the estimated probability field. Remember that these station values are truly superimposed on the estimated field. This is the reason that a station value of, say, 86 could lie between an 80 contour line and an 85 contour line. In contrast, the maps of estimated joint probability fields do not depict actual counted joint probability fields do not depict actual counted joint probability data at a station, but rather, joint probabilities estimated using the Boehm (1977) formula.

Figure A.1 is a map of the station locations in the demonstration area. The five-digit WMO number is plotted alongside the station location. These WMO numbers can be cross-referenced to Table A.1, which has other information pertaining to the station. Figure A.2 is also a map of the station locations, but the station elevations (in meters) are plotted. Figure A.3 is a map of this area's terrain based on the data in the USAFETAC 3DNEPH Terrain File. Figures A.4 through A.51 are maps of the actual climatological probabilities as analyzed by the Barnes algorithms for the months, time, and thresholds indicated.

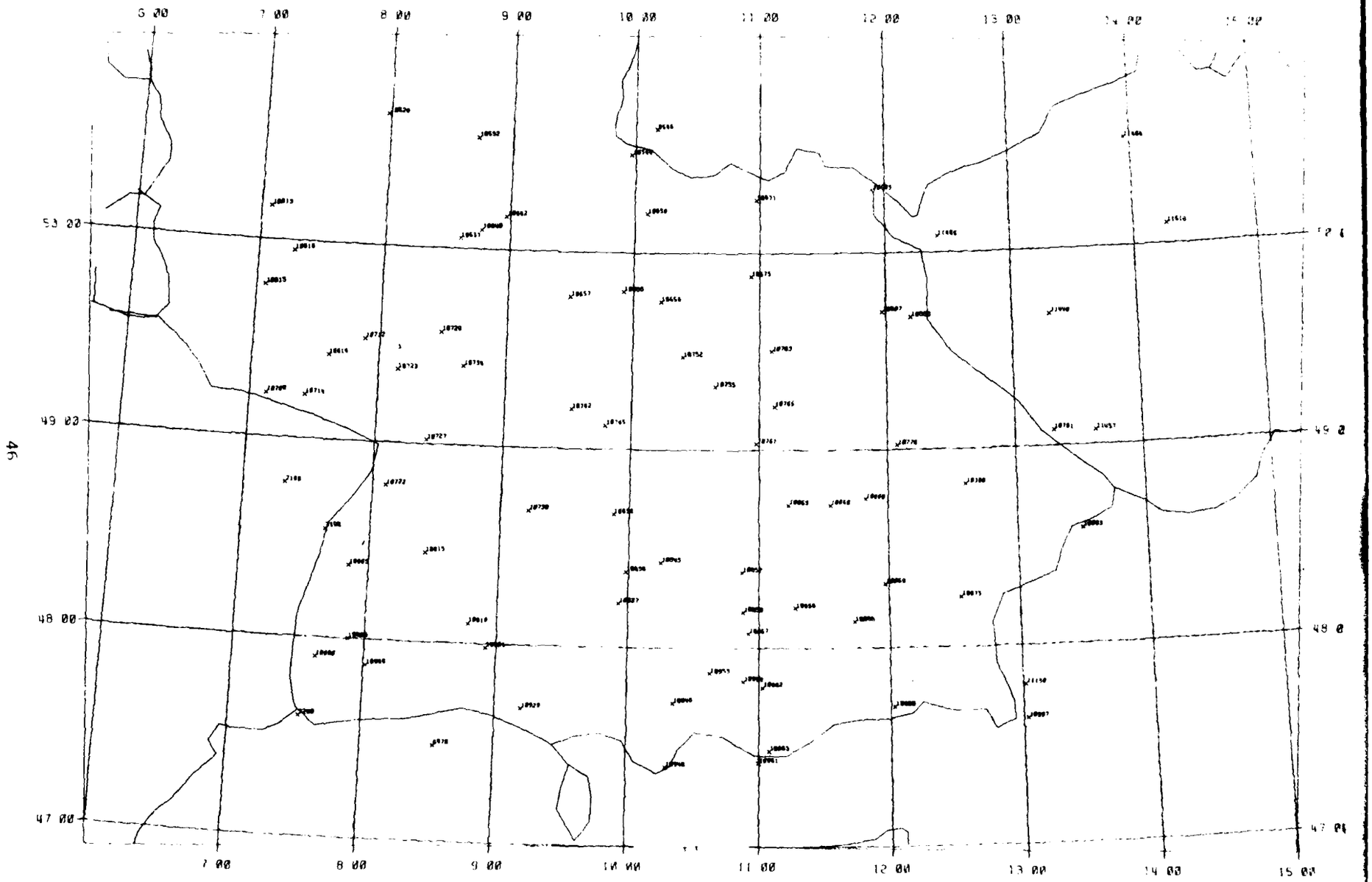


Figure A. Station WMO Numbers.

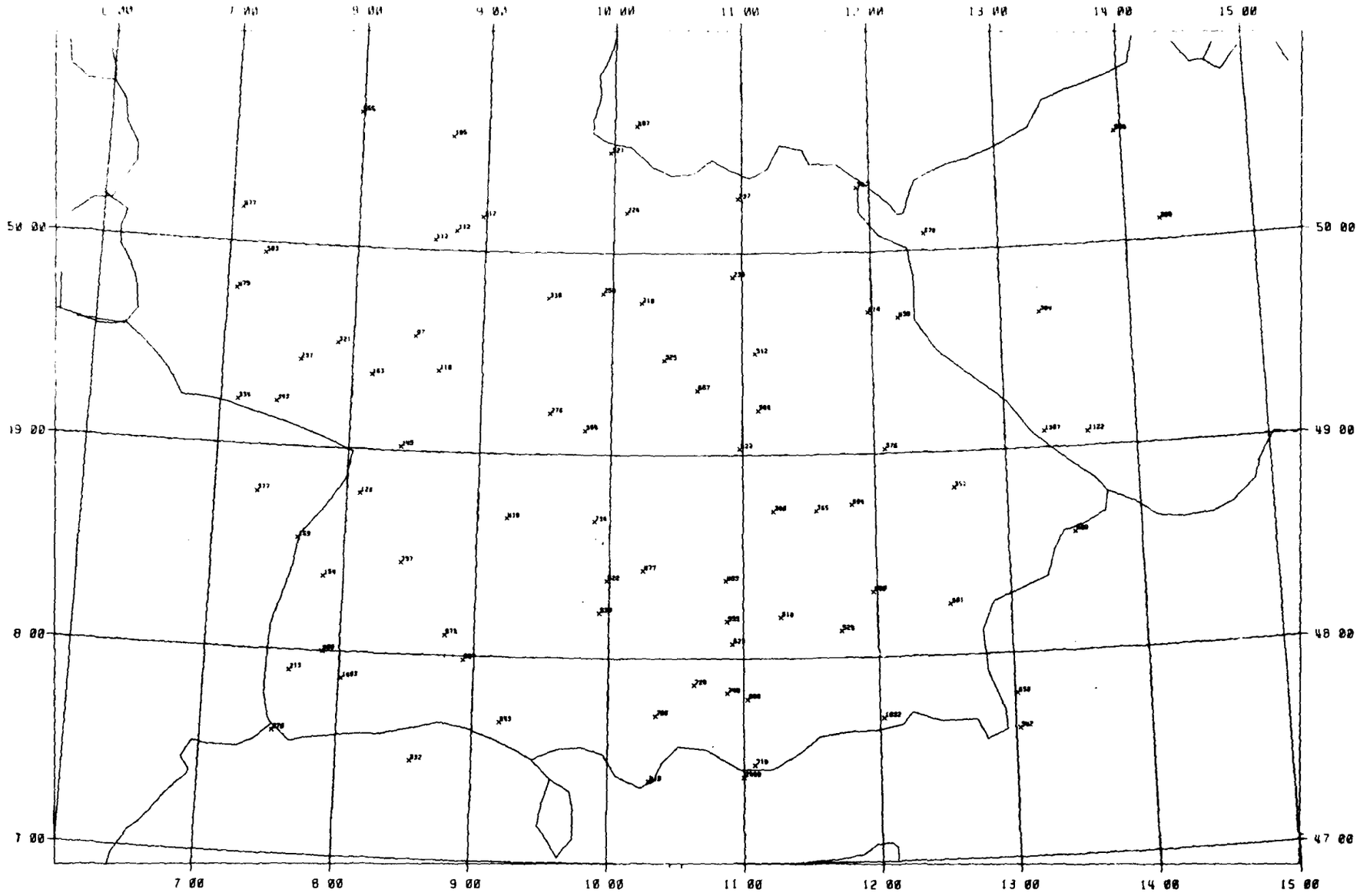


Figure A.2. Station Elevations.

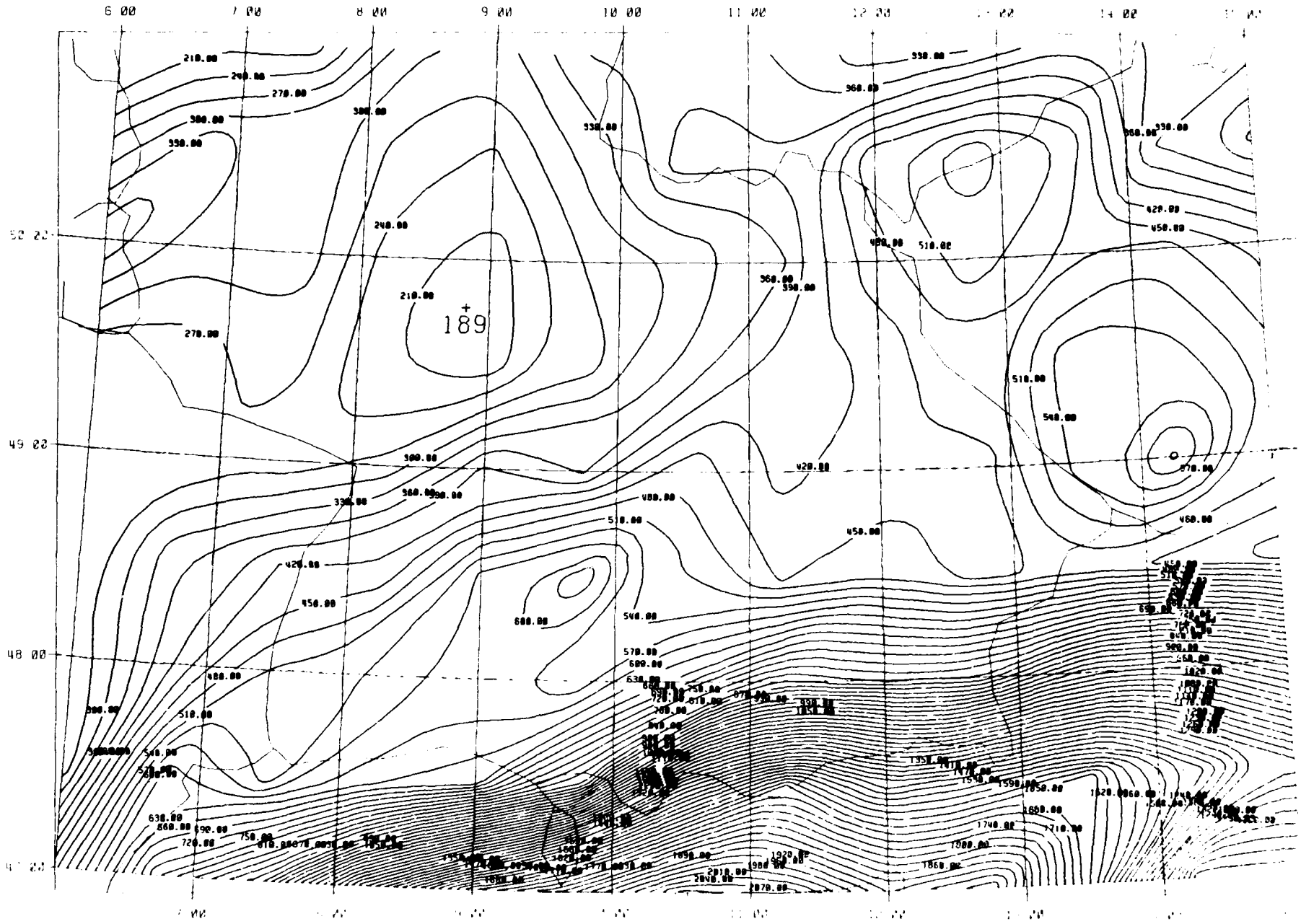


Figure A.3. Terrain Elevation (Meters). (Based on USAFETAC 30NEH Terrain File 1/8 mesh, Polar Stereographic projection).

A.2 Legend

Units--percent probability to the nearest percent

Projection--polar stereographic

Scale -Maps in this report are photo reductions of approximate scale 1:5,550,000. Original Versatec plots of Unconditional Probabilities are 1:1650000; Joint Probabilities are 1:1500000.

—80— -ADXOSCN generated contour.

×100 -WWMX generated latitude/longitude grid point estimate. Latitude grid interval = 0.5°, longitude grid interval = 0.75°.

×100 -Station data value

× -Missing station data

TABLE A.1. WMO Station Locations in the Demonstration Area.

WMO#	STATION	LATITUDE	LONG (E)	ELEVATION (M)
109540	Altenstadt, GER	47.83	10.88	740
107550	Ansbach/Kat., GER	49.32	10.65	467
108520	Augsburg, GER	48.38	10.87	463
72990	Bale/Mulhouse, FR	47.60	7.53	270
106750	Bamberg, GER	49.88	10.93	239
109970	Berchtesgadn, GER	47.63	13.03	542
109000	Bremgarten, GER	47.90	7.63	213
106130	Buchel/Cochem, GER	50.17	7.08	477
114060	Cheb, CZ	50.08	12.42	474
114570	Churanov, CZ	49.07	13.63	1122
106710	Coburg, GER	50.27	10.97	337
107295	Coleman, GER	49.57	8.48	97
106150	Deuselbach, GER	49.77	7.07	479
108690	Erding, GER	48.32	11.97	460
109080	Feldberg, GER	47.87	8.02	1493
108580	Ferstenelbrck, GER	48.20	11.28	518
106370	Frankfurt/M., GER	50.05	8.60	112
108030	Freiburg, GER	48.00	7.87	300
108150	Freudenstadt, GER	48.85	8.43	797
109630	Garmisch, GER	47.48	11.08	719
105320	Giessen, GER	50.57	8.72	195
106870	Grafenwohr, GER	49.70	11.97	414
107910	Grosser Falk, GER	49.08	13.30	1307
106160	Hahn, GER	49.95	7.28	503
106420	Hanau, GER	50.17	8.97	112
107340	Heidelberg, GER	49.40	8.67	110
106850	HOF, GER	50.32	11.90	567
109620	Hohenpeissenb, GER	47.80	11.03	986
107520	Illesheim, GER	49.47	10.40	325
108600	Ingolstaut, GER	48.72	11.55	365
95460	Kaltennordheim, GDR	50.63	10.17	487
107270	Karlsruhe, GER	49.02	8.40	145
109530	Kaufbeuren, GER	47.87	10.63	728
109460	Kempten, GER	47.72	10.35	705
106580	Kissinlen, GER	50.20	10.10	224
106590	Kitzingen, GER	49.75	10.22	210

108180	Klippeneck, GER	48.10	8.78	973
109290	Konstanz, GER	47.68	9.20	443
108050	Lahr, GER	48.37	7.85	154
108570	Landsberg, GER	48.07	10.92	623
108370	Laudheim, GER	48.22	9.93	538
108560	Lechfeld, GER	48.18	10.88	555
108450	Leipheim, GER	48.43	10.25	477
105260	Marienberg, GER	50.67	7.98	555
106400	Maurice Rose, GER	50.10	8.77	112
114640	Mileskova, CZ	50.55	13.95	836
108750	Muhdorf, GER	48.25	12.55	401
108660	Munchen/R., GER	48.13	11.73	529
108530	Neuberg, GER	48.72	11.23	380
109210	Neuhausen, GER	47.98	8.92	807
107230	Neustadt/WF, GER	49.37	8.15	163
107630	Nurnburg, GER	49.50	11.12	312
109480	Oberstdorf, GER	47.40	10.30	810
107420	Ohringen, GER	49.20	9.53	276
108930	Passau, GER	48.58	13.50	409
71860	Phalsbourg, FR	48.77	7.32	377
114480	Plezen/Dobra, CZ	49.67	13.30	364
115180	Prague/Ruzyne, CZ	50.10	14.27	369
106140	Ramstein, GER	49.43	7.60	237
107760	Regensburg, GER	49.02	12.08	376
107650	Roth, GER	49.22	11.12	386
107080	Saarbrucken, GER	49.22	7.13	334
111500	Salzburg, AUS	47.80	13.02	450
107450	Schwabisch H., GER	49.12	9.80	398
107120	Sembach, GER	49.52	7.88	321
108605	Siegenburg, GER	48.75	11.82	404
107220	Sollingen, GER	48.78	8.10	123
107880	Staubing, GER	48.82	12.60	352
108360	Stotten, GER	48.67	9.88	734
71900	Strasbourg/Entzh, FR	48.55	7.65	153
107380	Stuttgart/Echter	48.68	9.22	419
108380	Ulm, GER	48.38	9.98	522
105440	Wasserkuppe, GER	50.50	9.97	921
106880	Weiden, GER	49.67	12.20	438
107610	Weissenburg, GER	49.03	10.98	422
109800	Wendelstein, GER	47.70	12.03	1832
106570	Wertheim, GER	49.77	9.50	338
106550	Wurzburg, GER	49.80	9.92	259
109610	Zugspitze, GER	47.42	11.00	2960
66700	Zurica/Kloten, SW	47.48	8.55	432
107140	Zweibrucken, GER	49.22	7.43	343

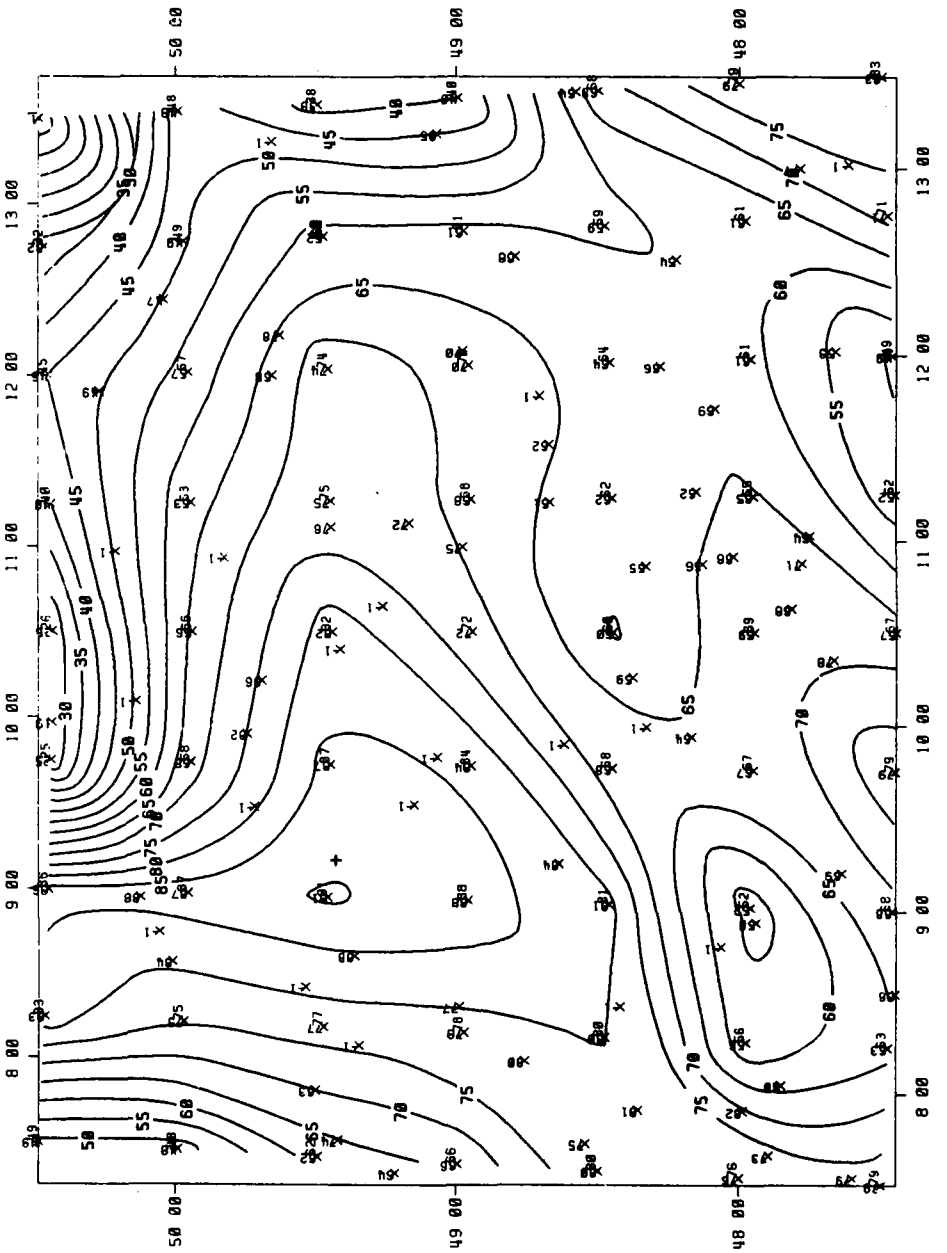


Figure A.4. Estimated Probability Fields, January, 02-04Z, Ceiling .GT. 1000FT AGL. (23 missing observations.)

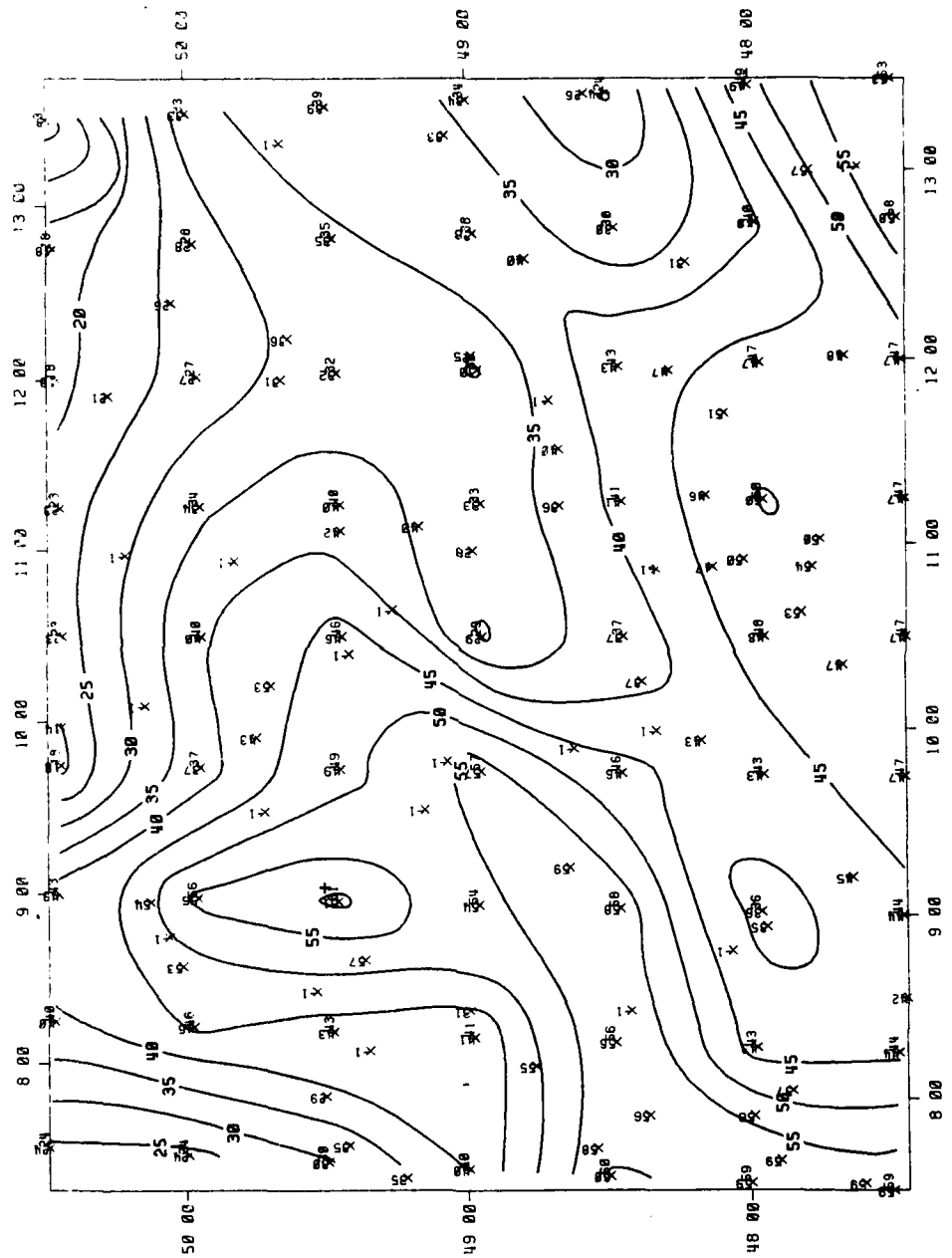


Figure A.5. Estimated Probability Fields, January, 02-04Z, Ceiling .GT. 3000FT AGL. (23 missing observations.)

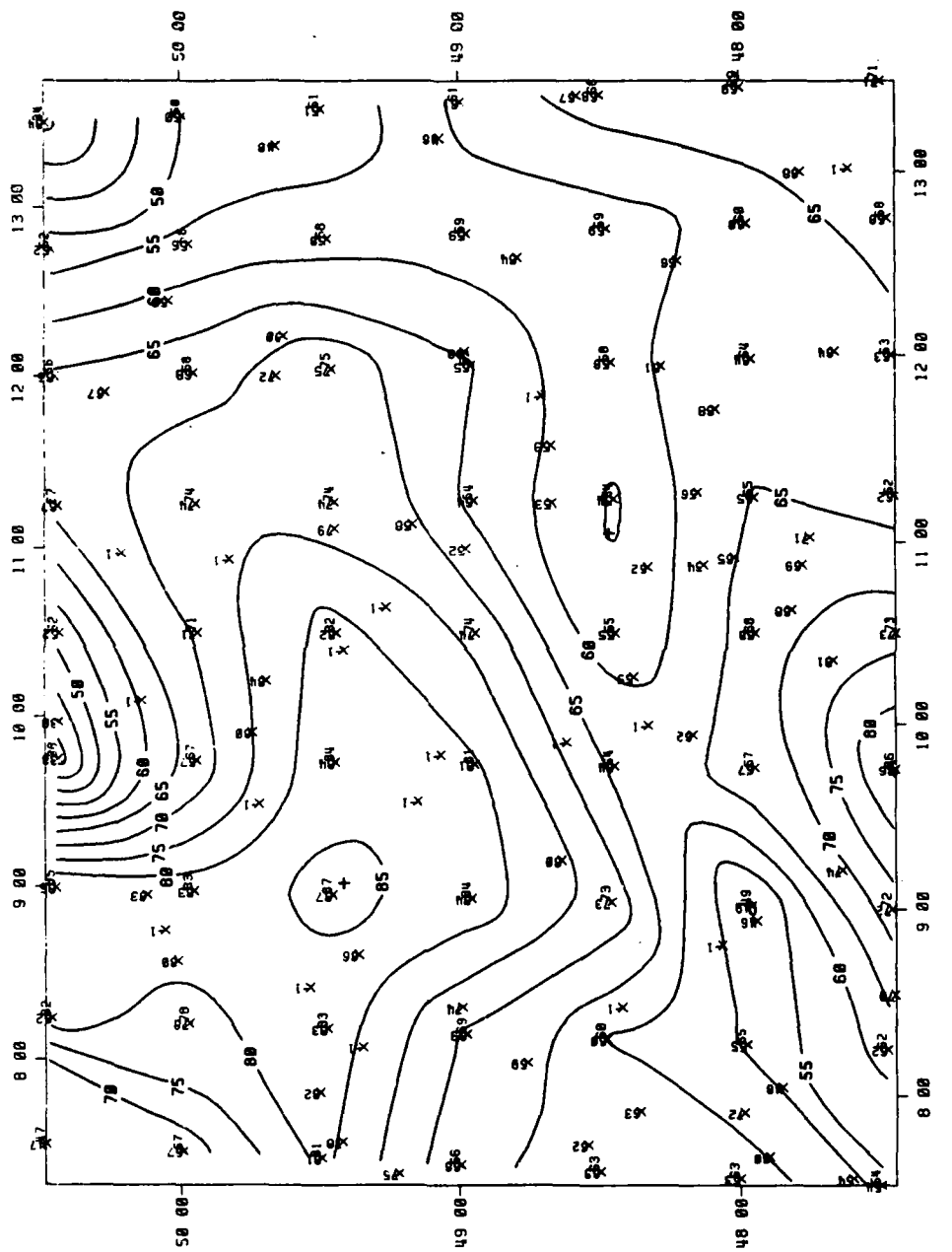


Figure A.6. Estimated Probability Fields, January, 02-04Z, Visibility .GT. 2MI. (21 missing observations.)

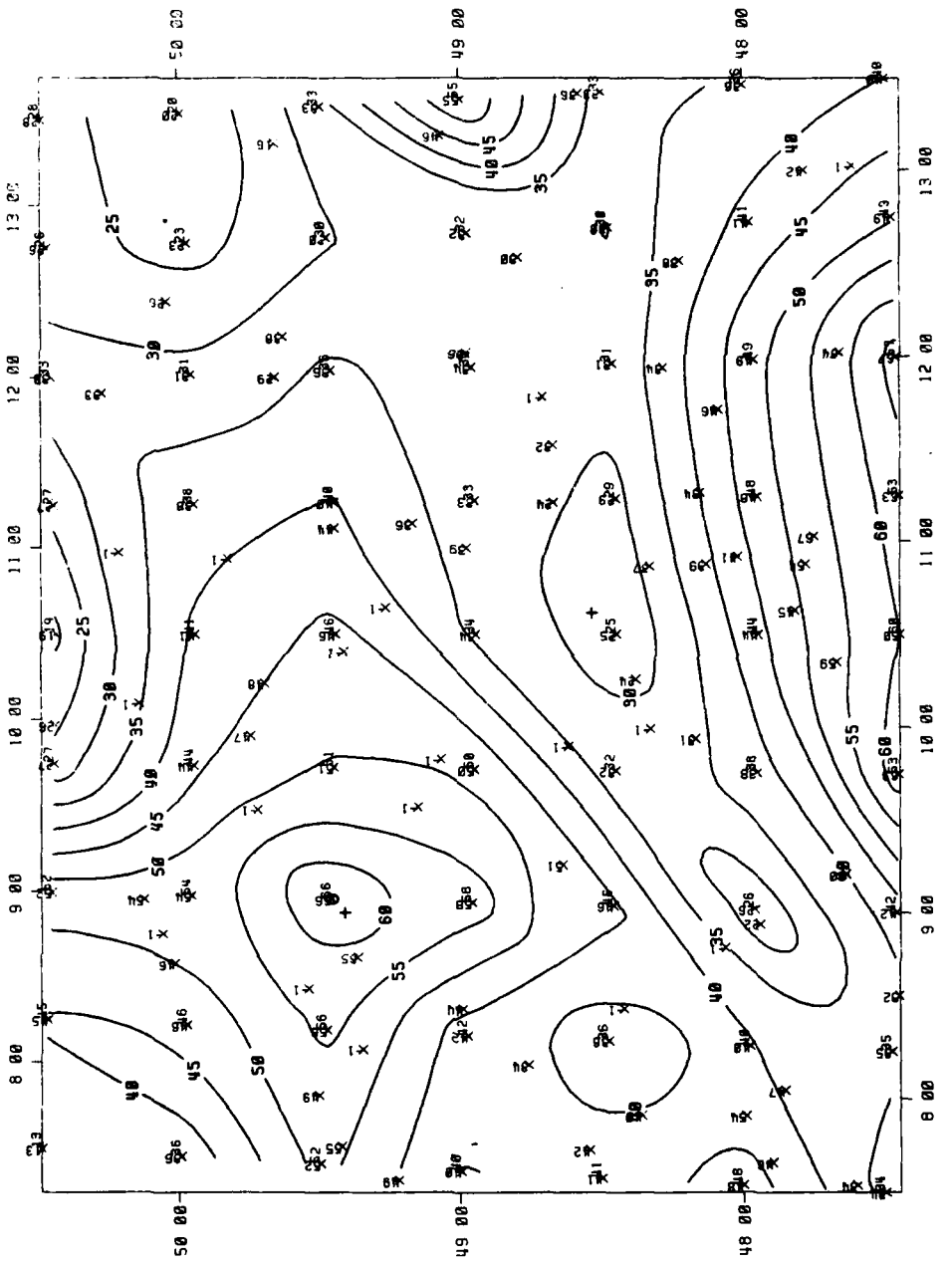


Figure A.7. Estimated Probability Fields, January, 02-04Z, Visibility .GT. 5MI. (21 missing observations.)

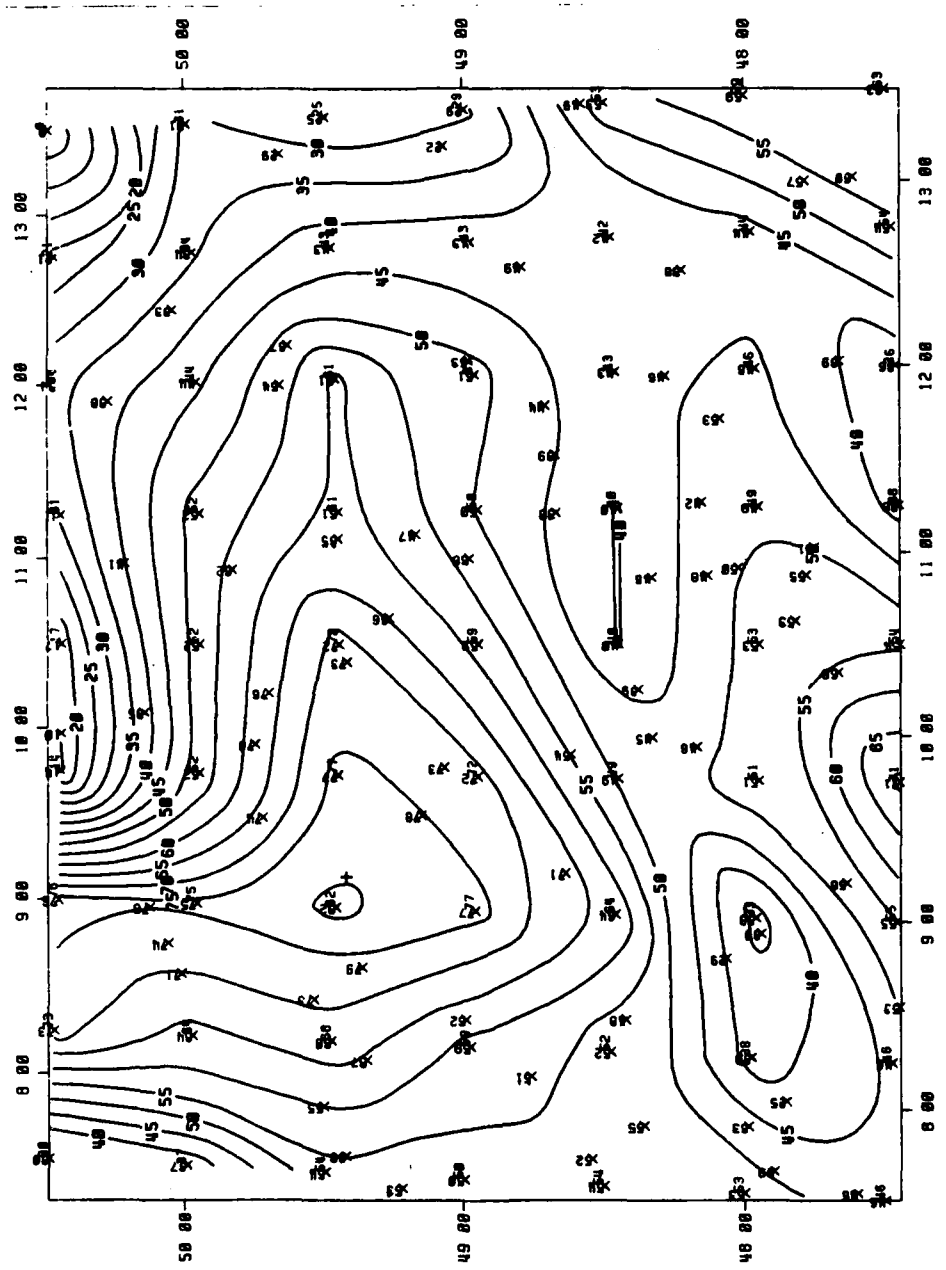


Figure A.8. Estimated Probability Fields,
January, 02-04Z, Joint CIG/VIS .LT. 1000/2.

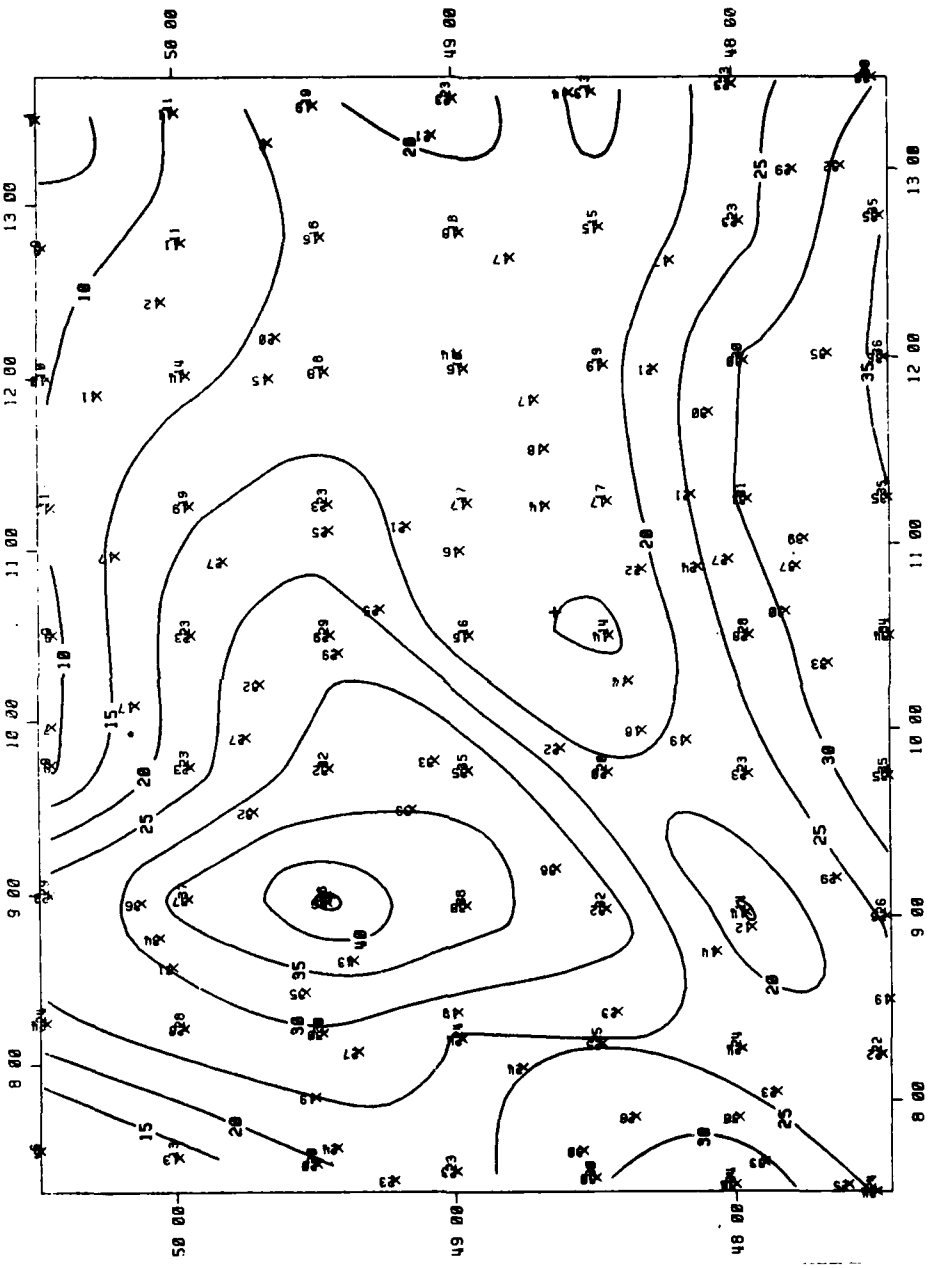


Figure A.9. Estimated Probability Fields, January, 02-04Z, Joint CIG/VIS .LT. 3000/5.

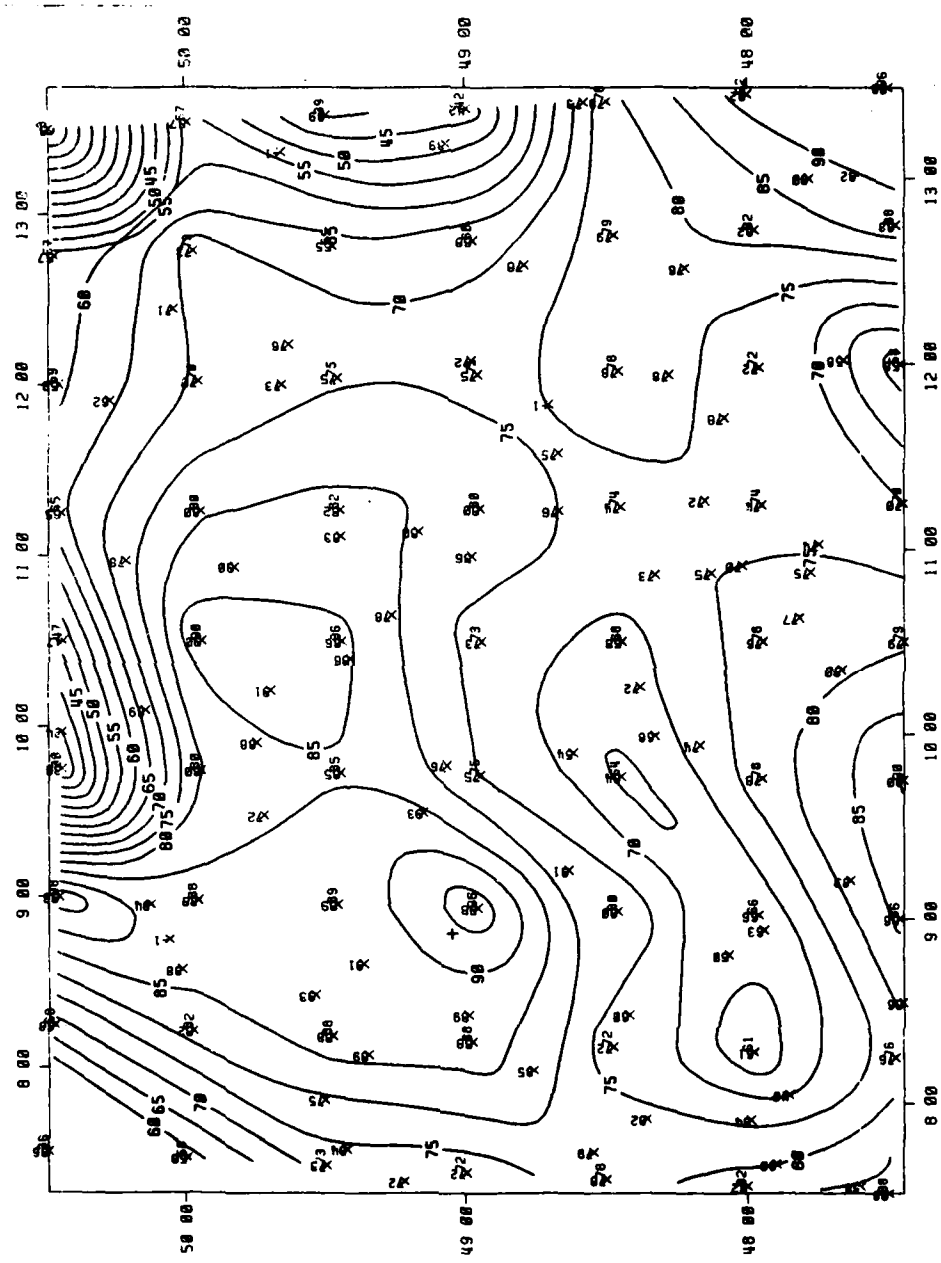


Figure A.10. Estimated Probability Fields,
January 14-16Z, Ceiling .GT. 1000FT AGL.
(4 missing observations.)

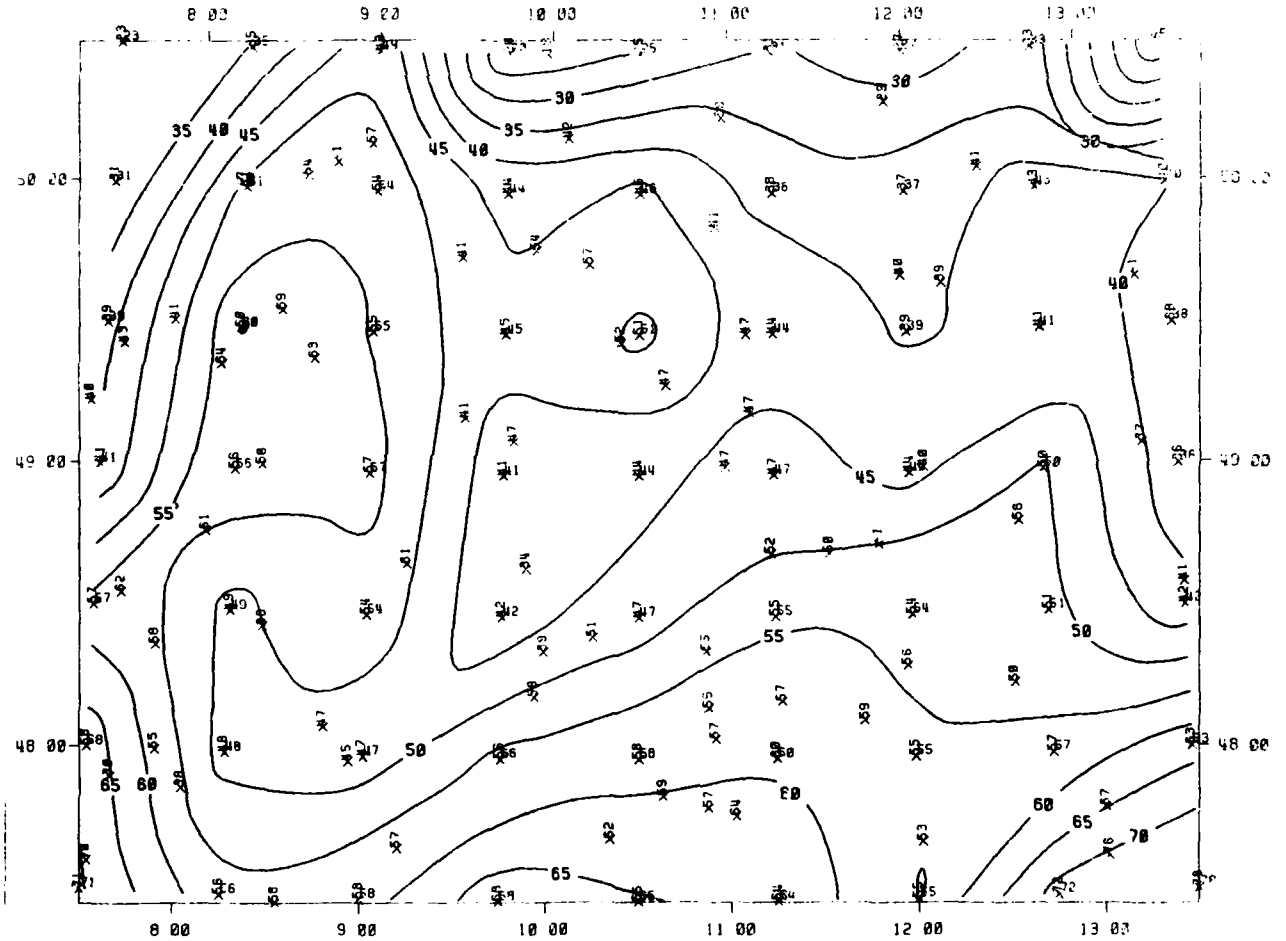


Figure A.11. Estimated Probability Fields, January, 14-16Z, Ceiling .GT. 3000FT AGL. (4 missing observations.)

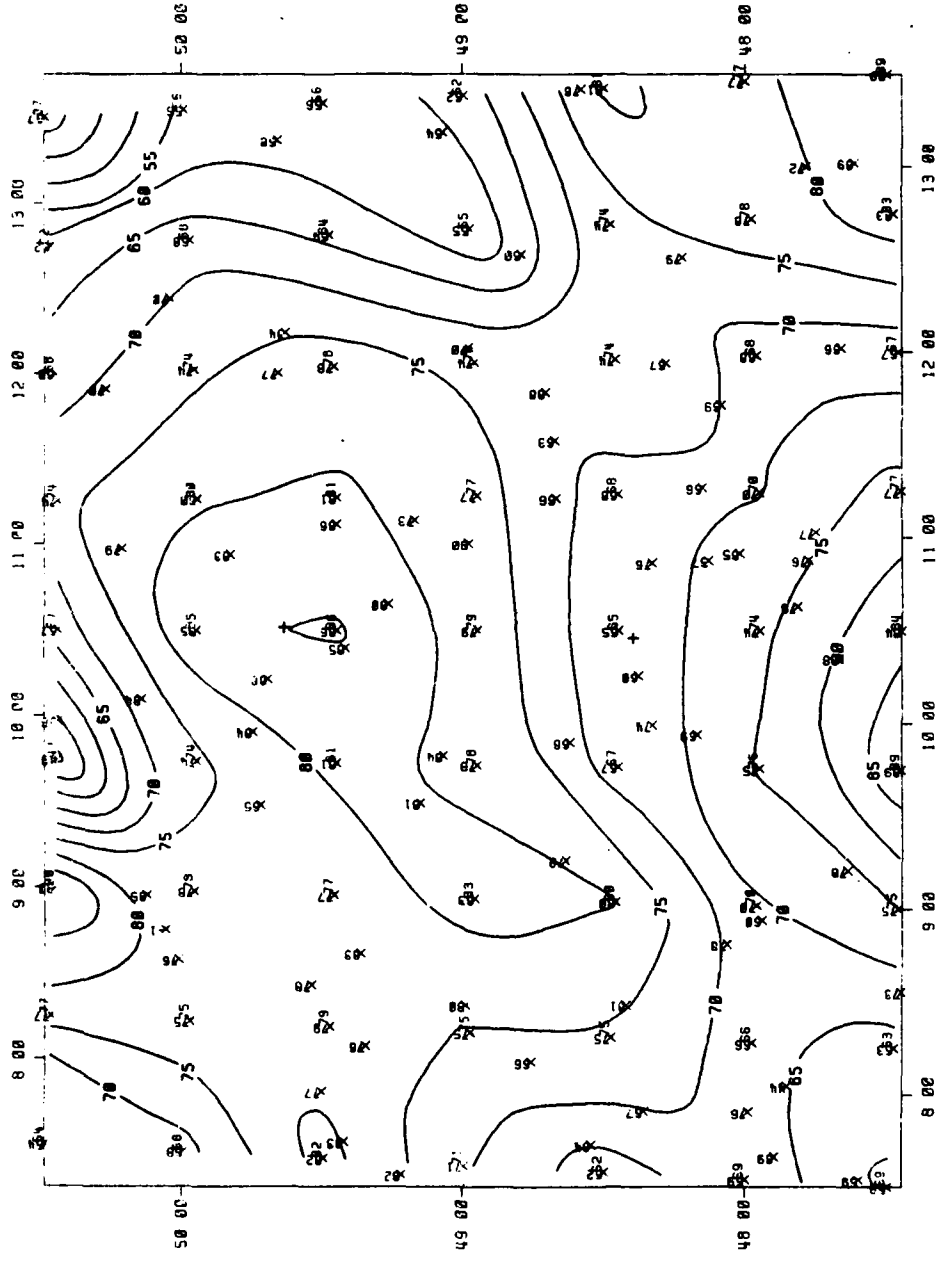


Figure A.12. Estimated Probability Fields, January, 14-16Z, Visibility .GT. 2MI. (1 missing observation.)

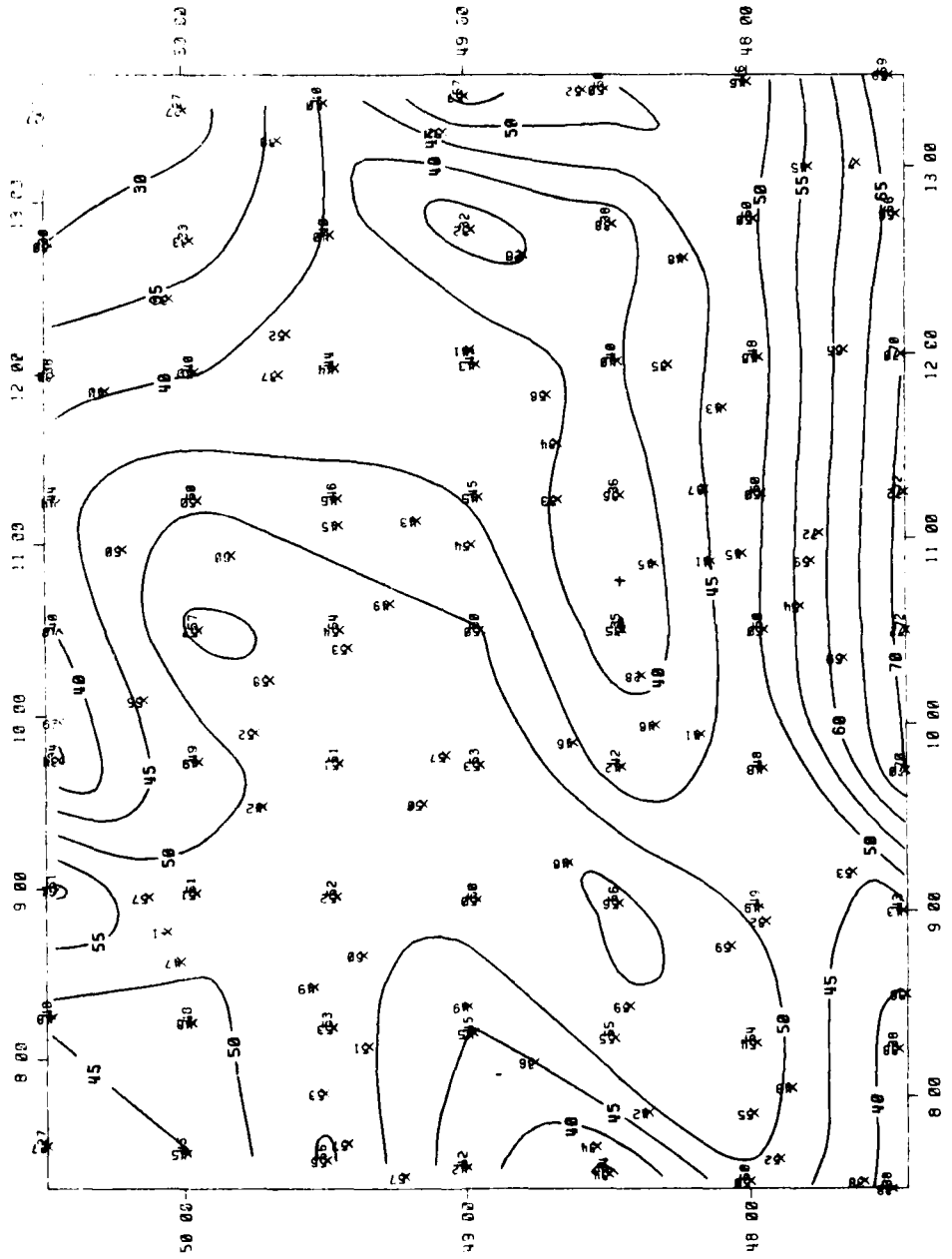


Figure A.13. Estimated Probability Fields, January, 14-16Z, Visibility .GT. 5MI. (1 missing observation.)

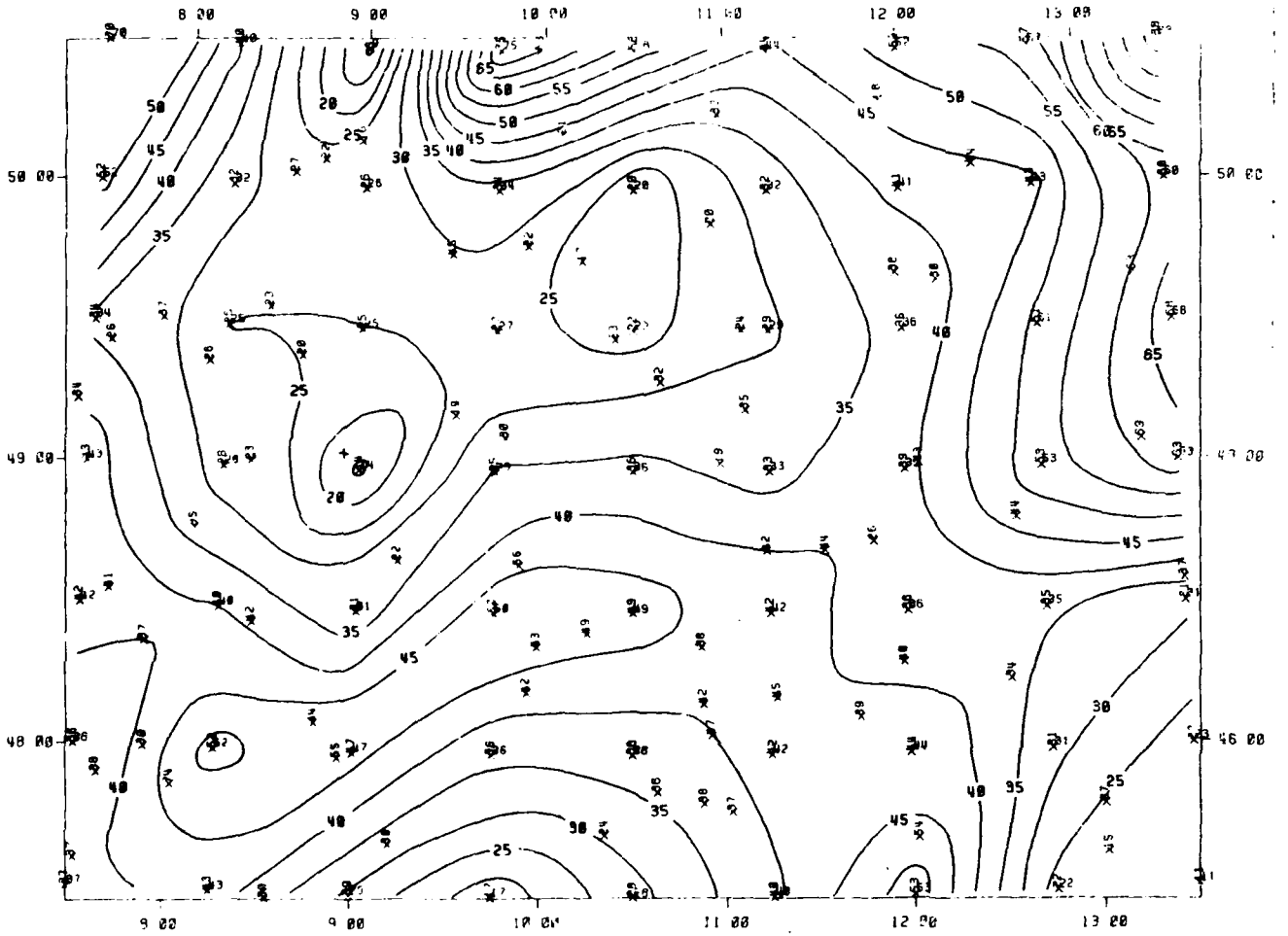


Figure A.14. Estimated Probability Fields, January, 14-16Z, Joint CIG/VIS .LT. 1000/2.

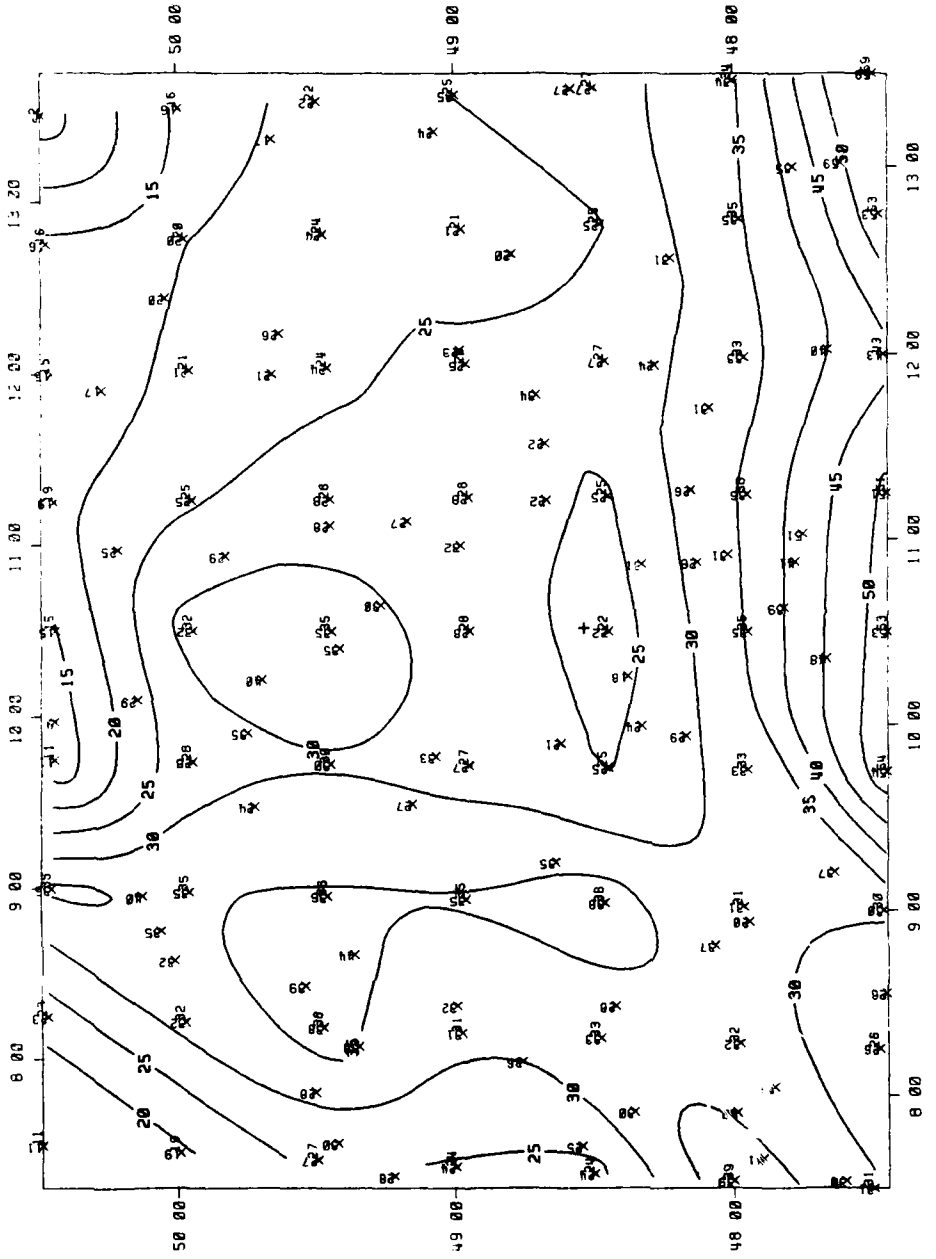


Figure A.15. Estimated Probability Fields, January, 14-16Z, Joint CIG/VIG .LT. 3000/5.

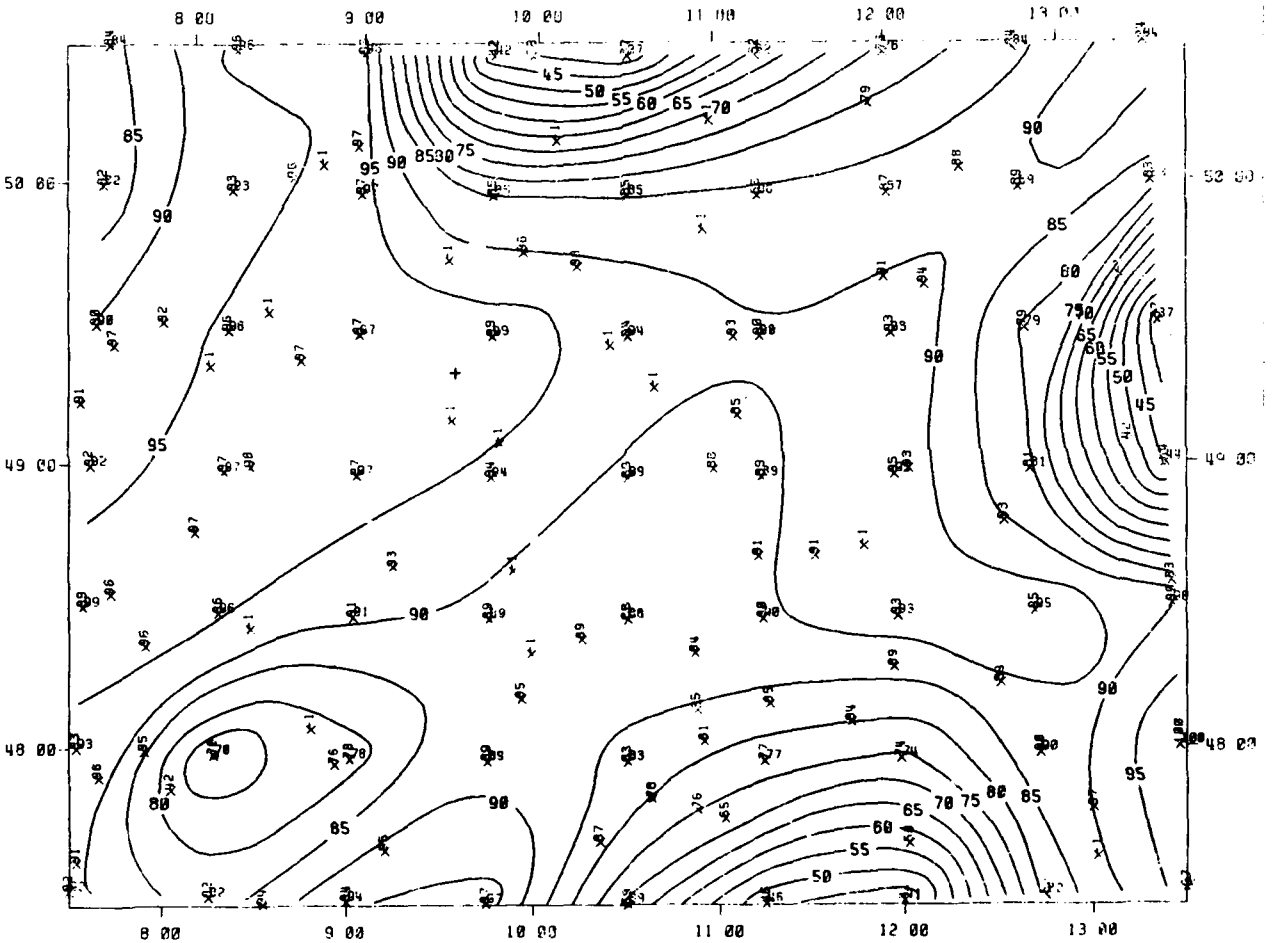


Figure A.16. Estimated Probability Fields,
April, 02-04Z, Ceiling .GT. 1000FT AGL.
(25 missing observations.)

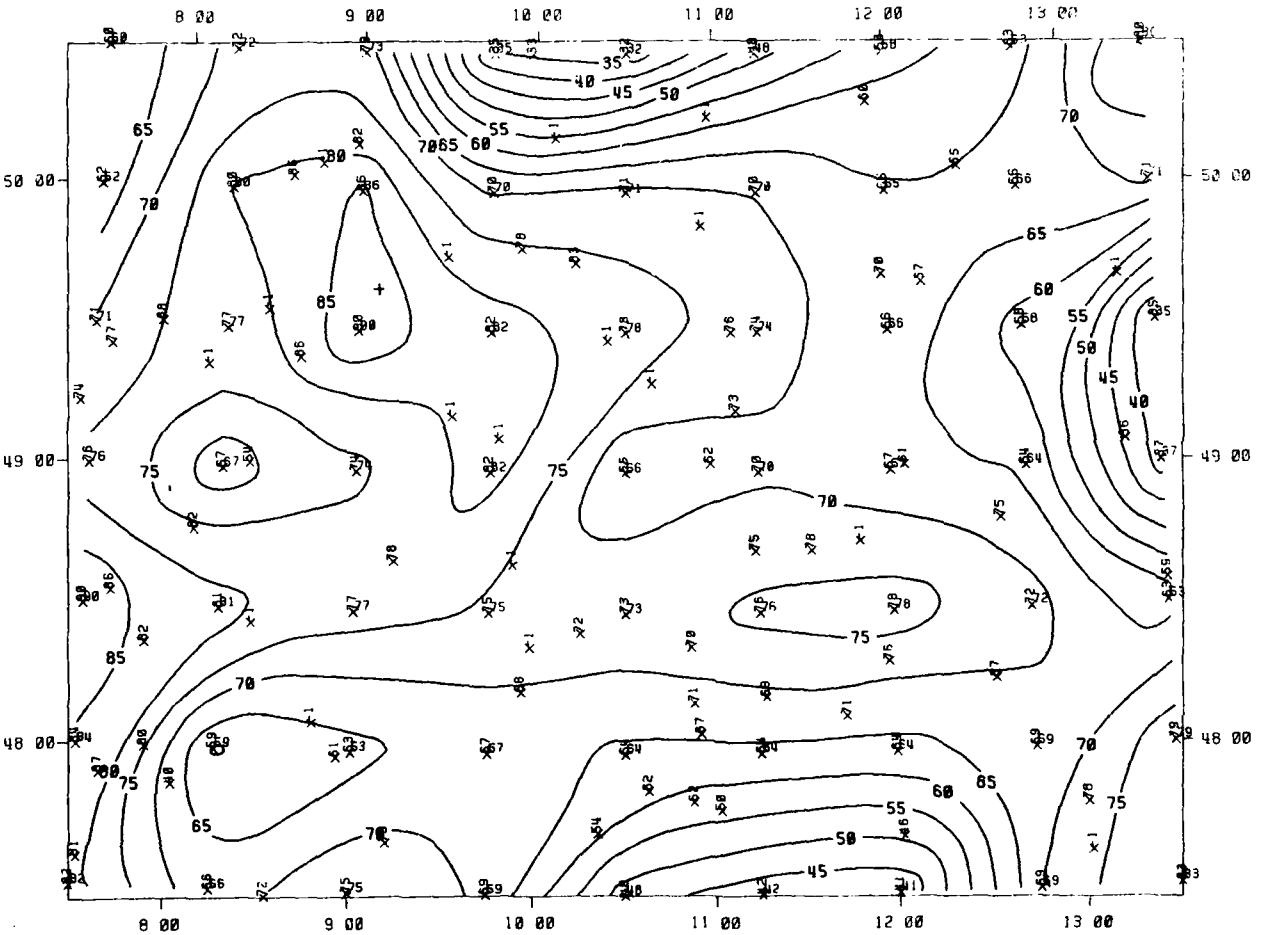


Figure A.17. Estimated Probability Fields,
April, 02-04Z, Ceiling .GT. 3000FT AGL.
(25 missing observations.)

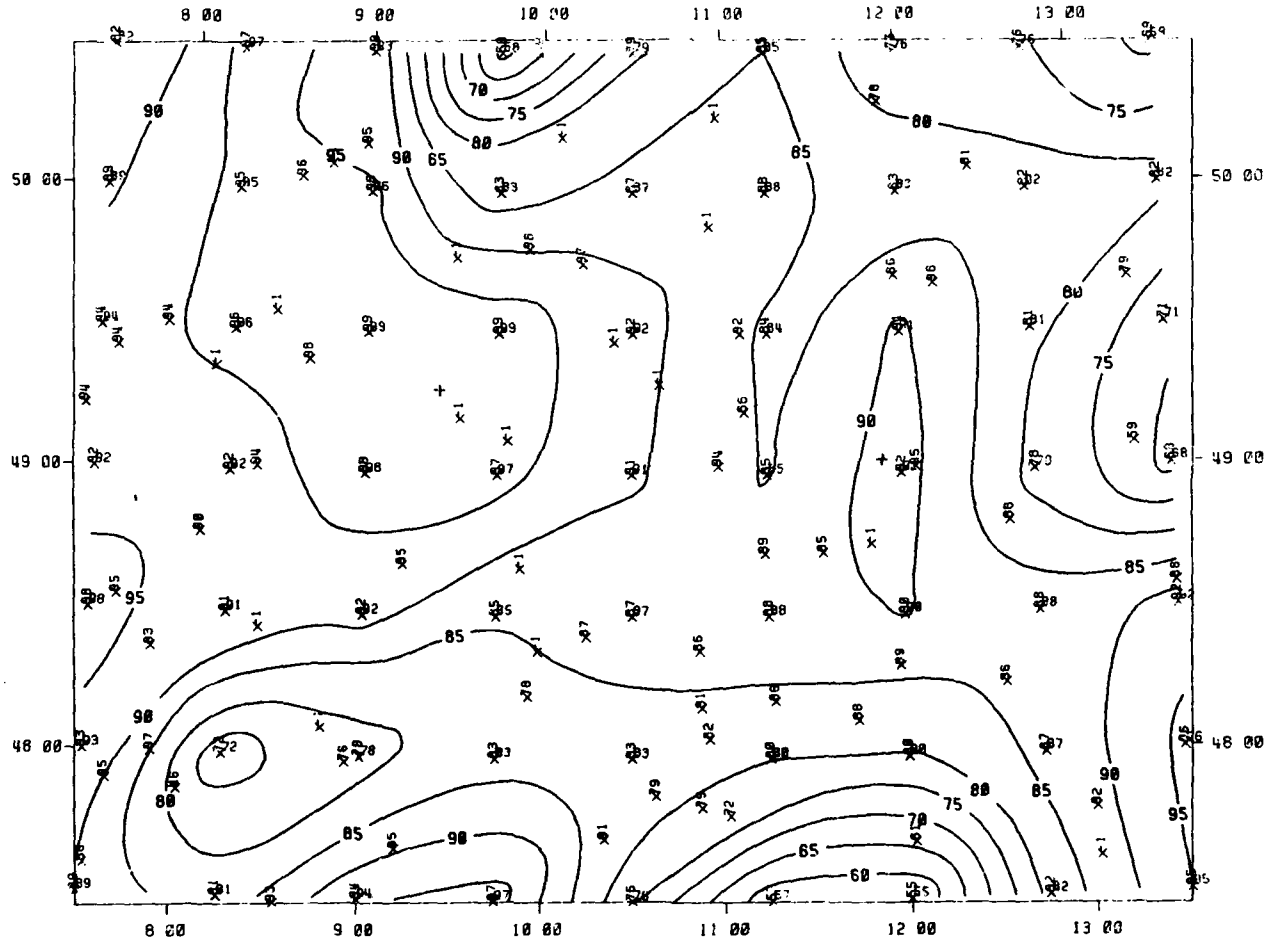


Figure A.18. Estimated Probability Fields, April, 02-04Z, Visibility .GT. 2MI. (21 missing observations.)

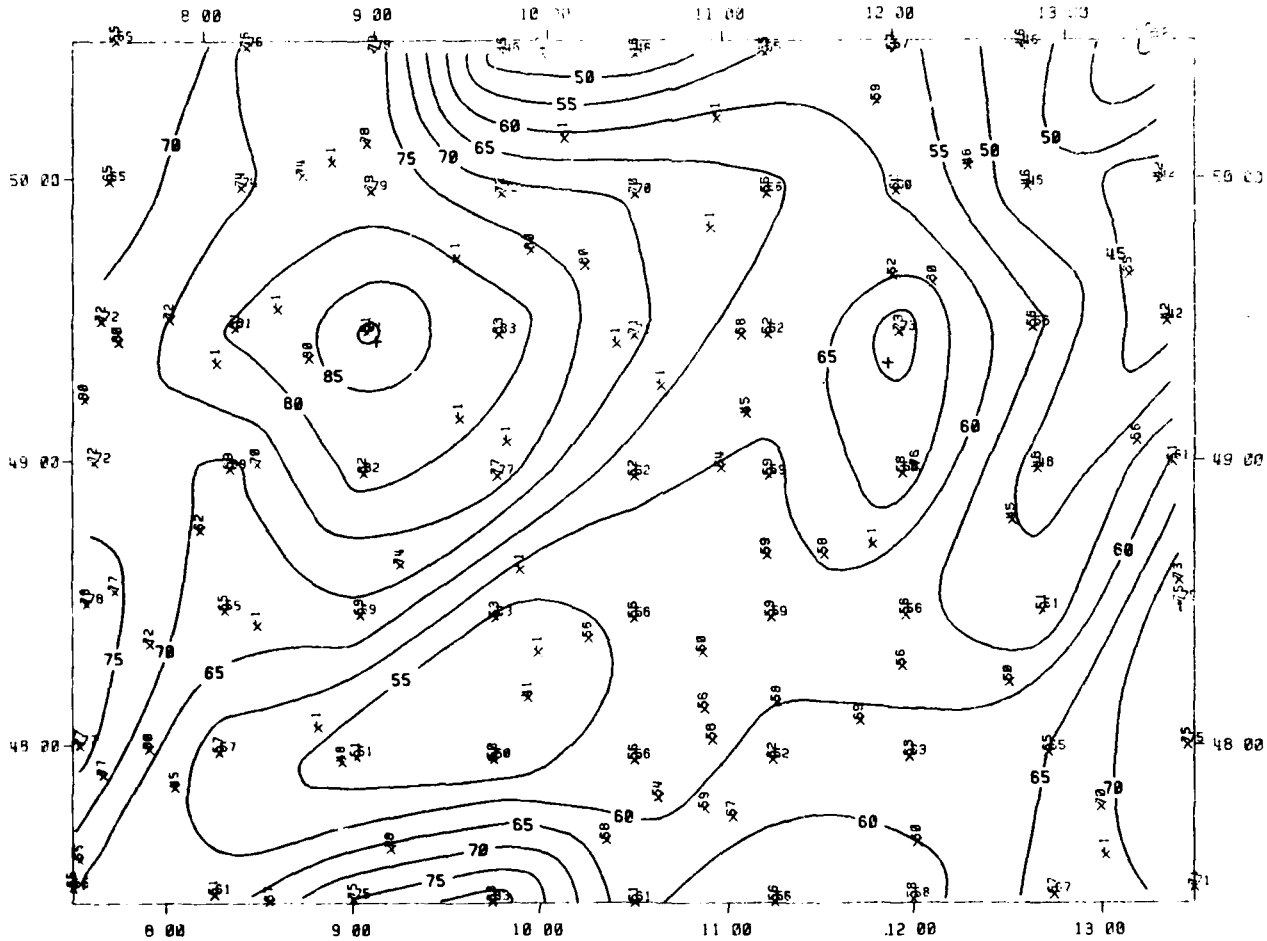


Figure A.19. Estimated Probability Fields, April, 02-04Z, Visibility .GT. 5MI. (21 missing observations.)

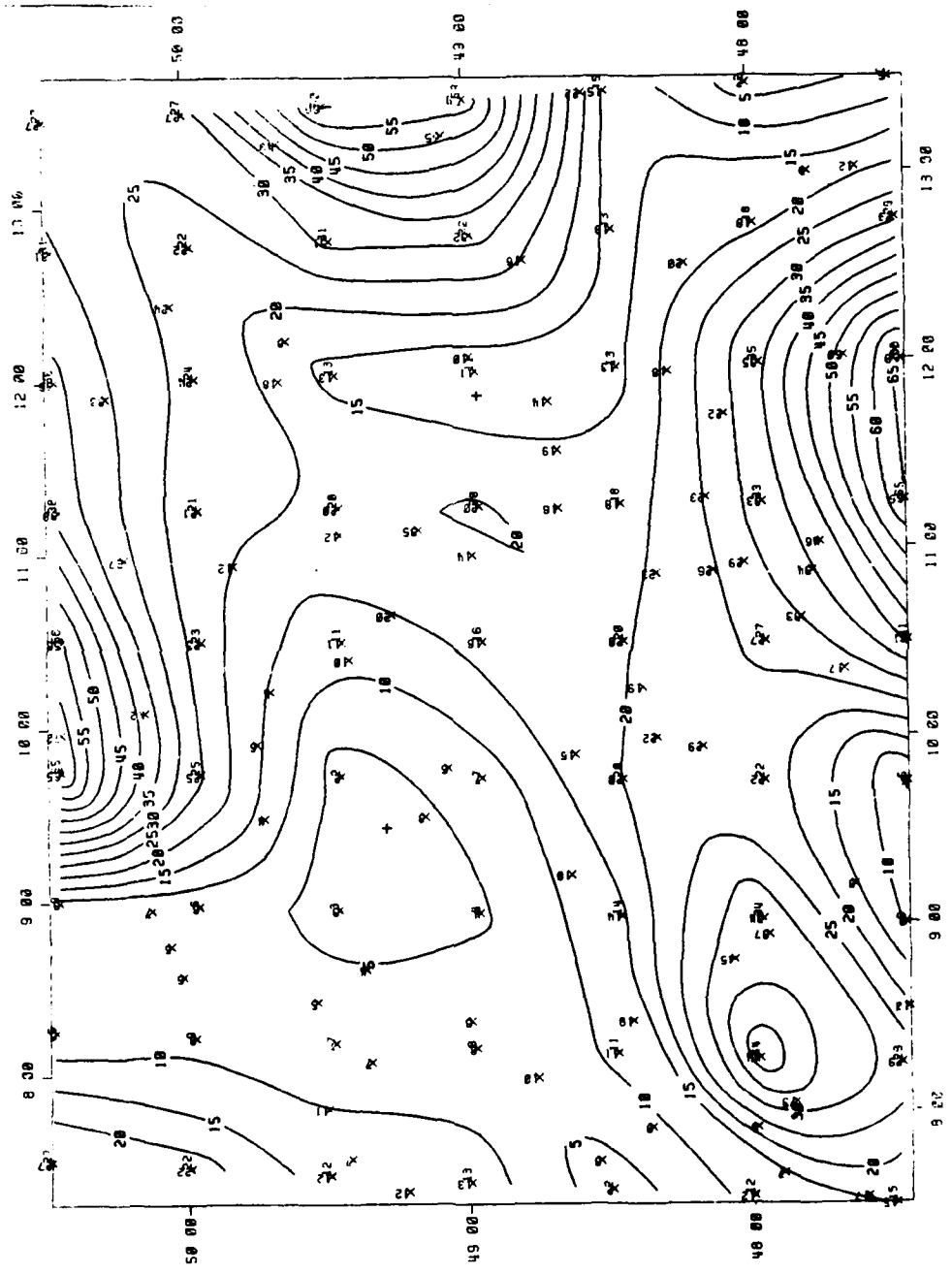


Figure A.20. Estimated Probability Fields, April, 02-04Z, Joint CIG/VIS .LT. 1000/2.

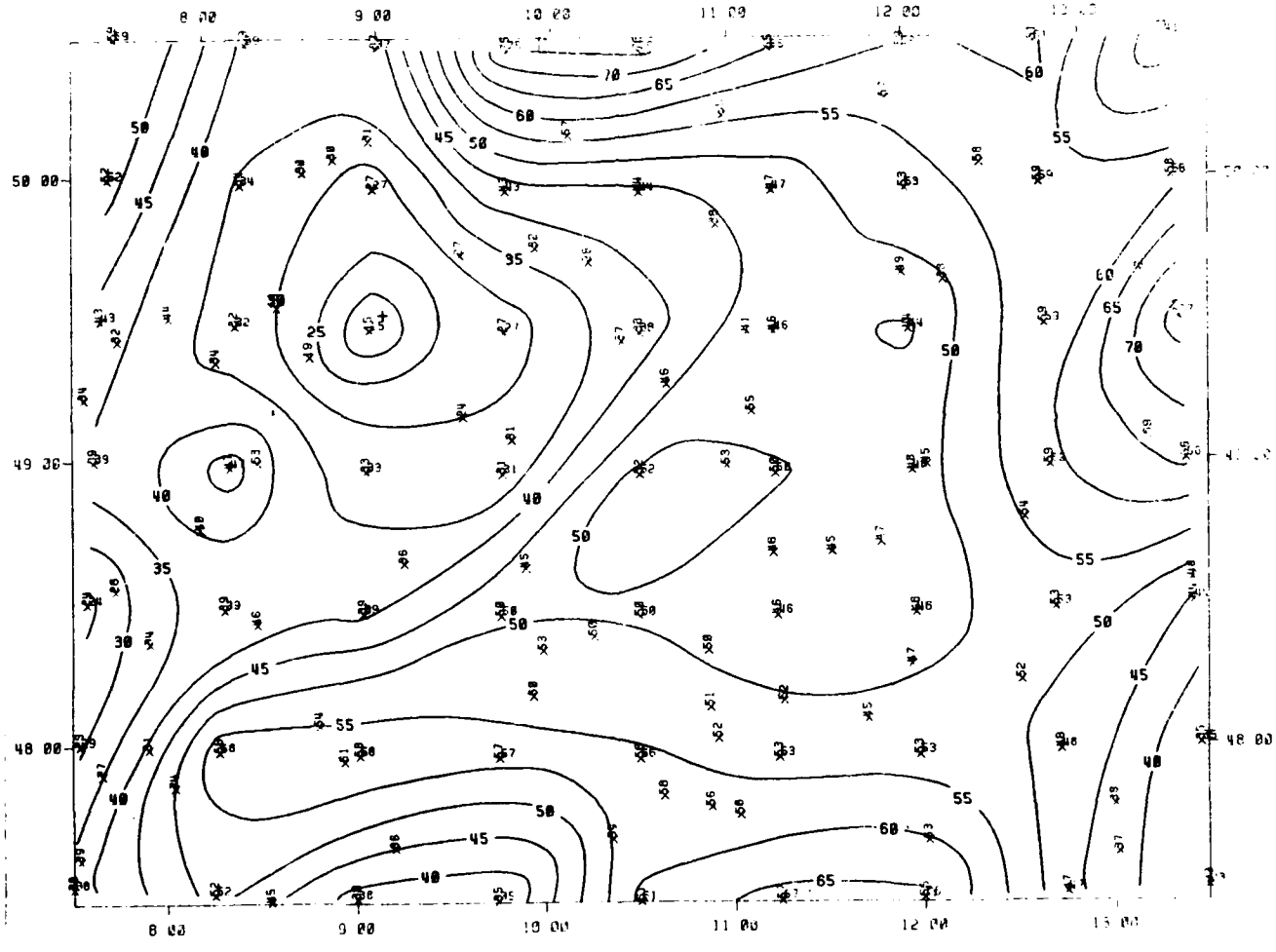


Figure A.21. Estimated Probability Fields, April, 02-04Z, Joint CIG/VIS .LT. 3000/5.

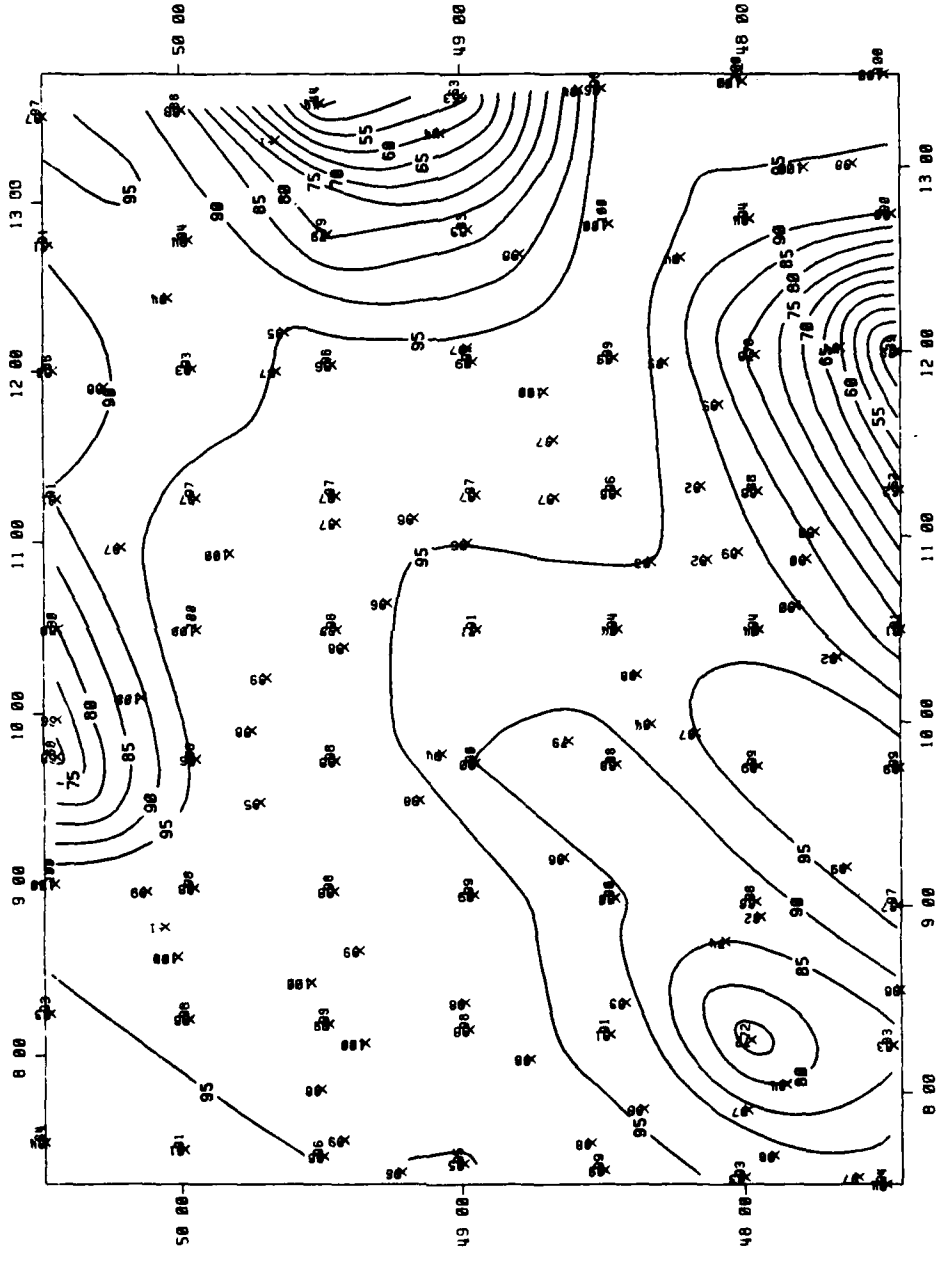


Figure A.22. Estimated Probability Fields,
April, 14-16Z, Ceiling .GT. 1000FT AGL.
(4 missing observations.)

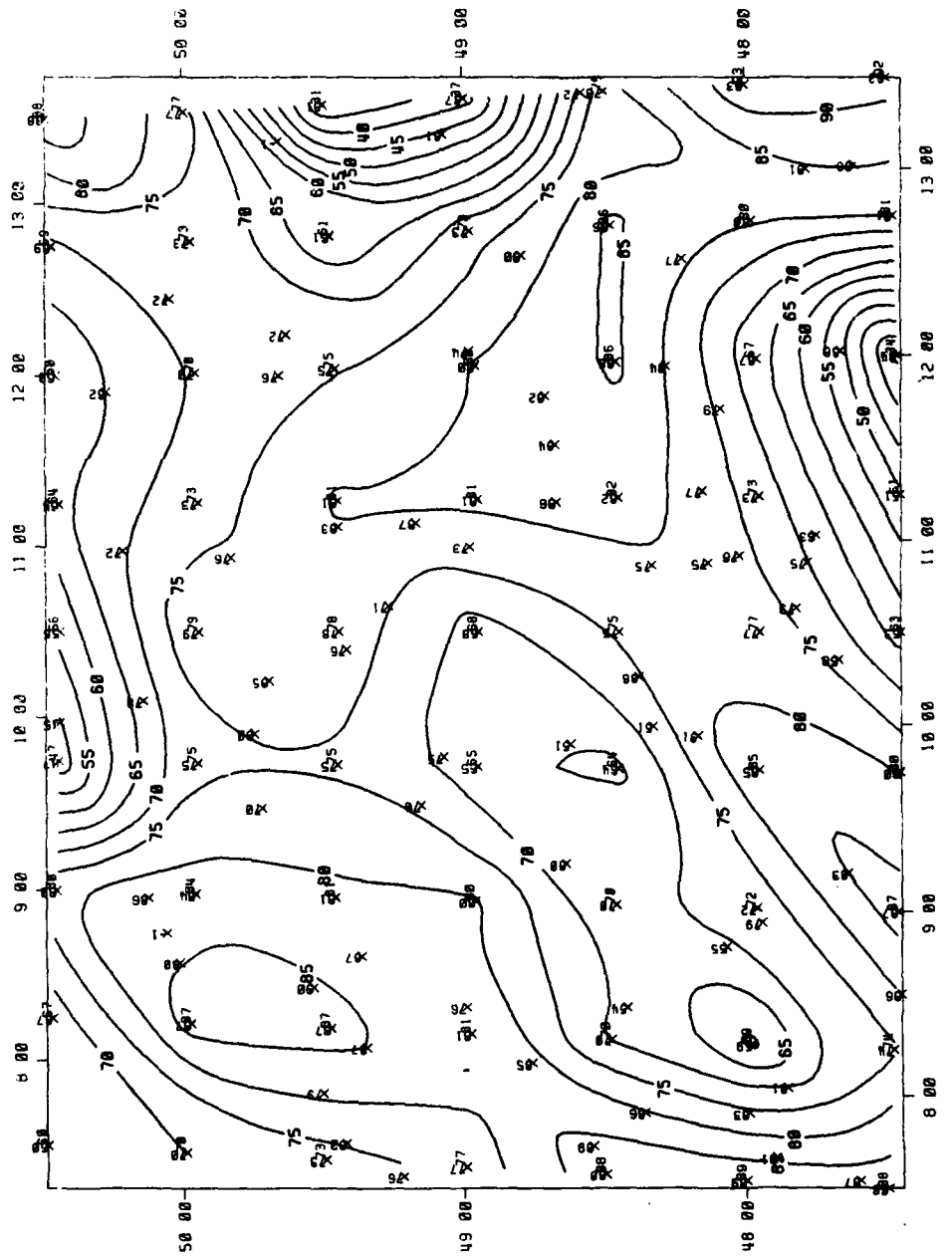


Figure A.23. Estimated Probability Fields, April, 14-16Z, Ceiling .GT. 3000FT AGL. (4 missing observations.)

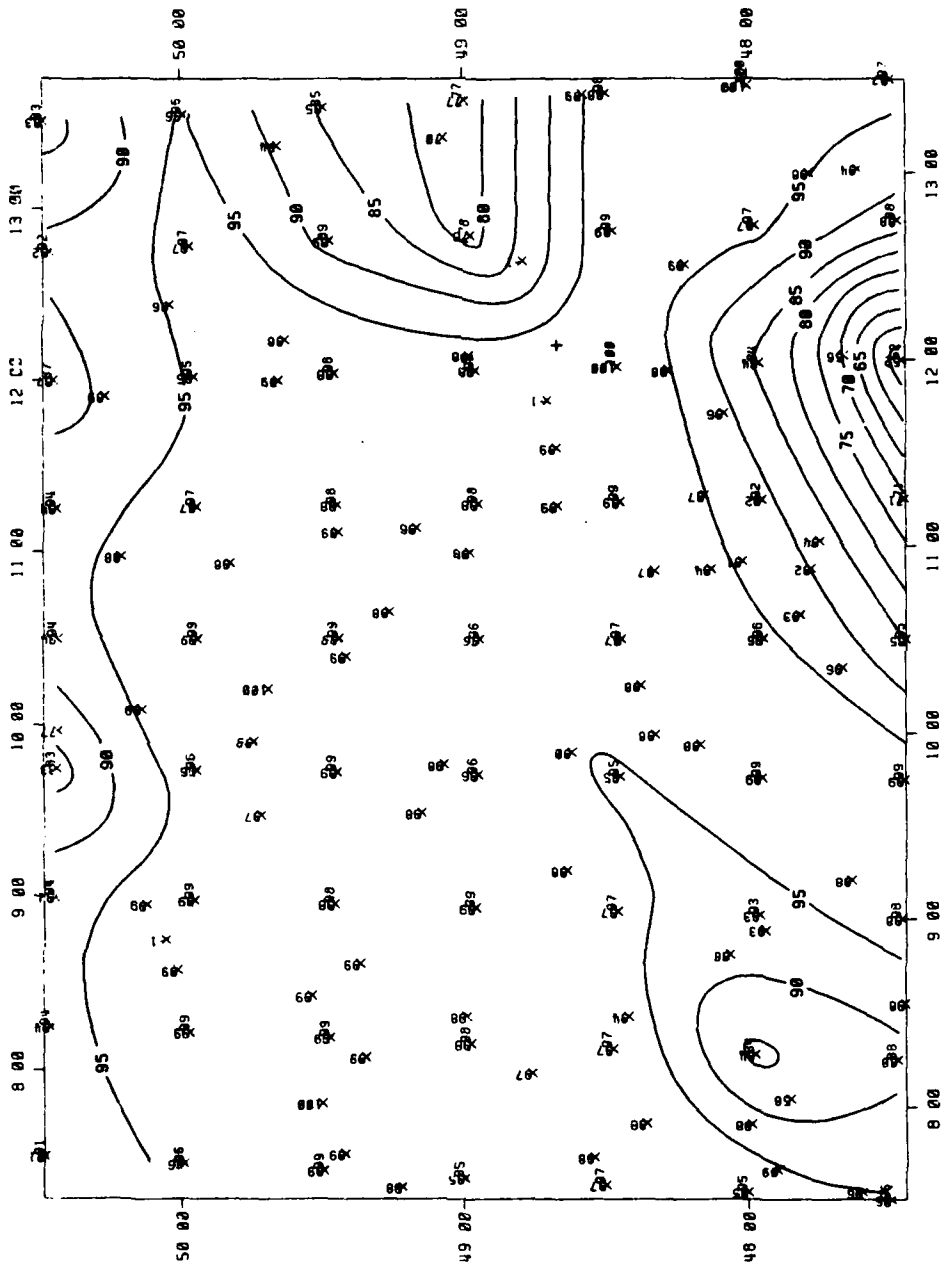


Figure A.24. Estimated Probability Fields, April, 14-16Z, Visibility .GT. 2MI. (4 missing observations.)

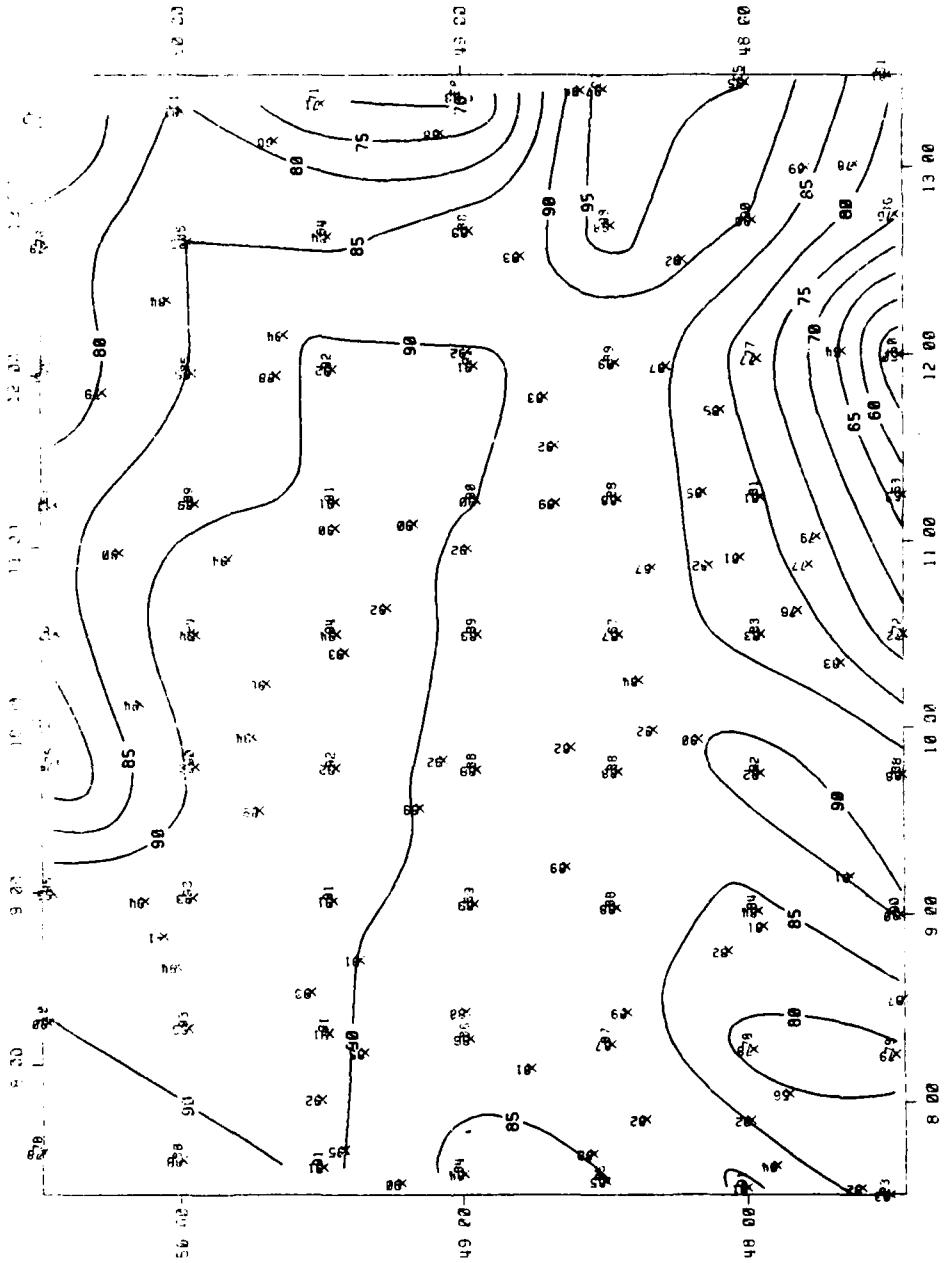


Figure A.25. Estimated Probability Fields,
April, 14-16Z, Visibility .GT. 5MI.
(1 missing observation.)

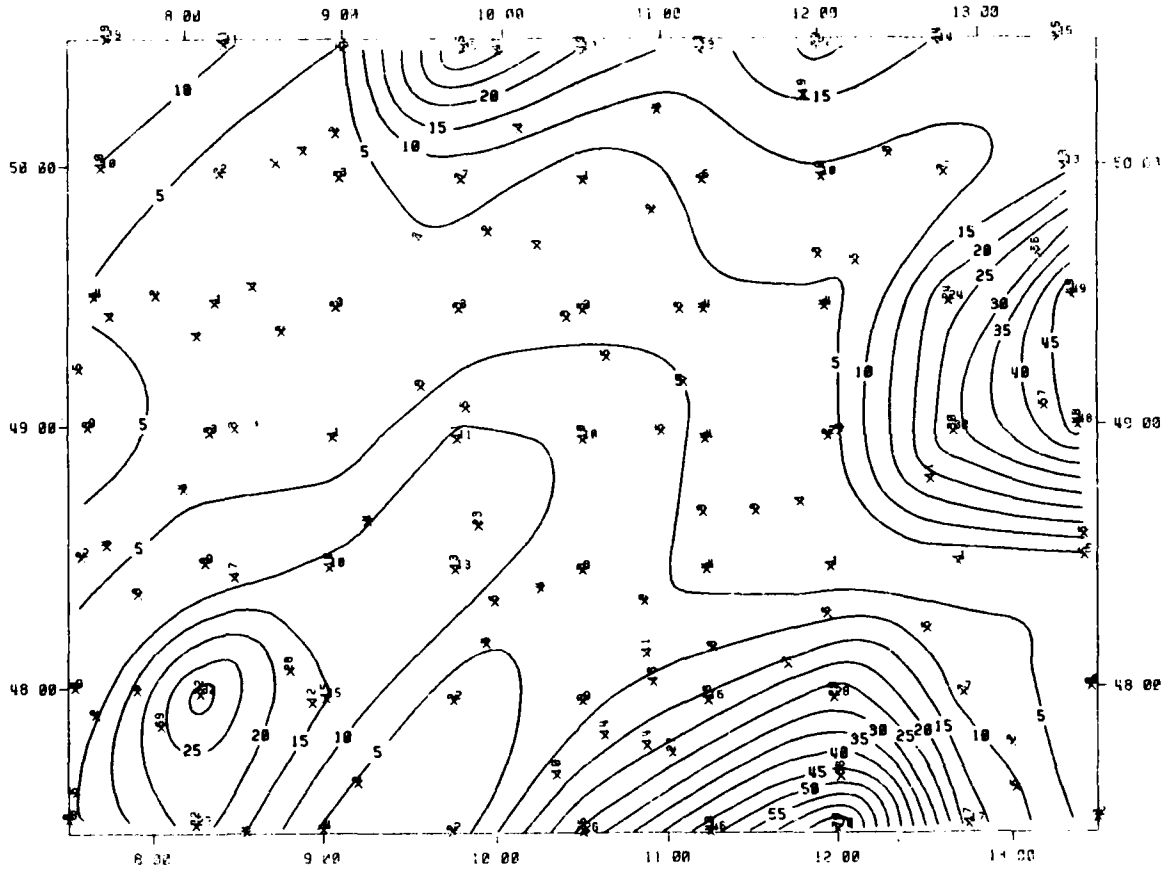


Figure A.26. Estimated Probability Fields, April, 14-16Z, Joint CIG/VIS .LT. 1000/2.

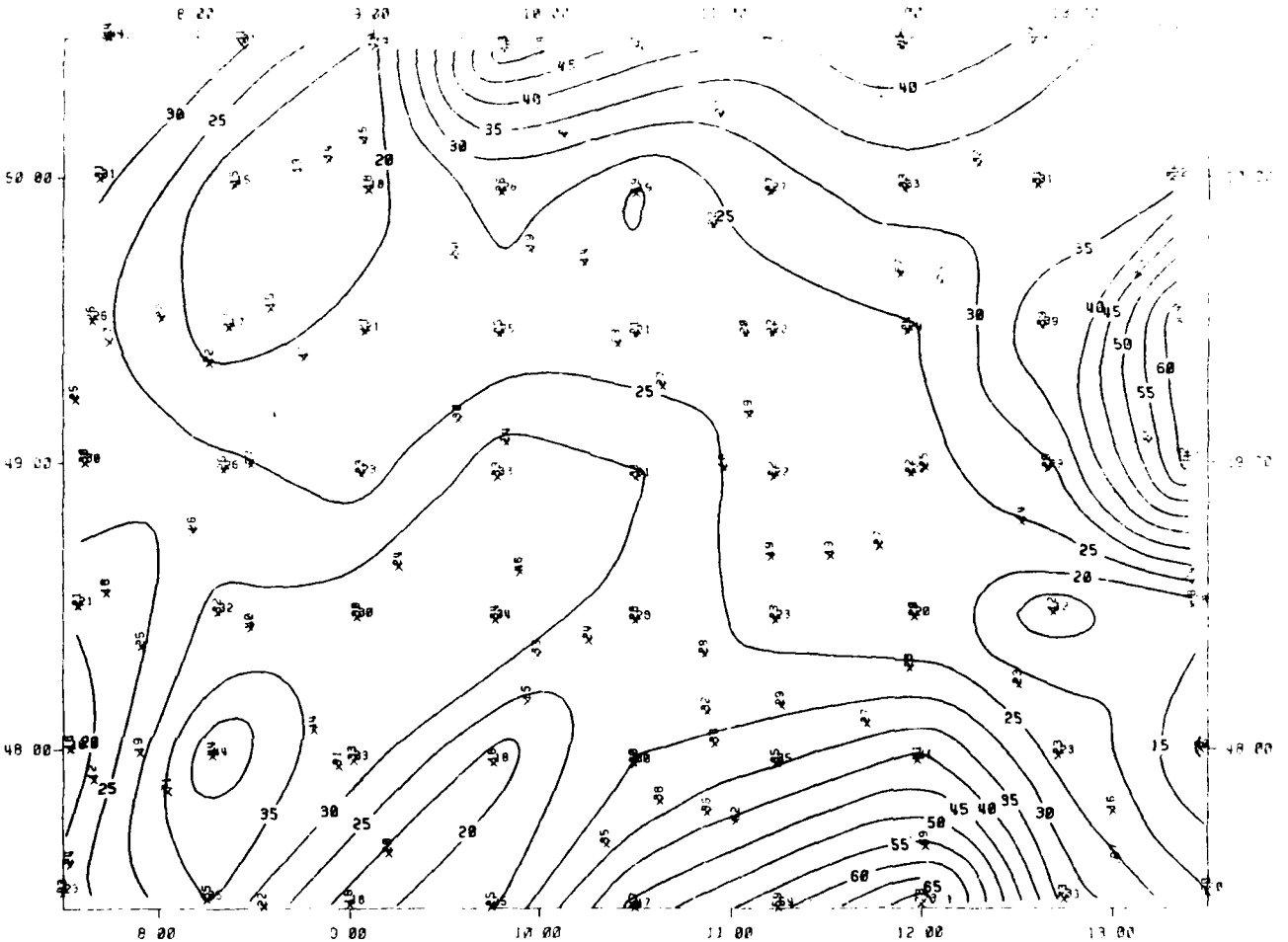


Figure A.27. Estimated Probability Fields, April, 14-16Z, Joint CIG/VIS .LT. 3000/5MI.

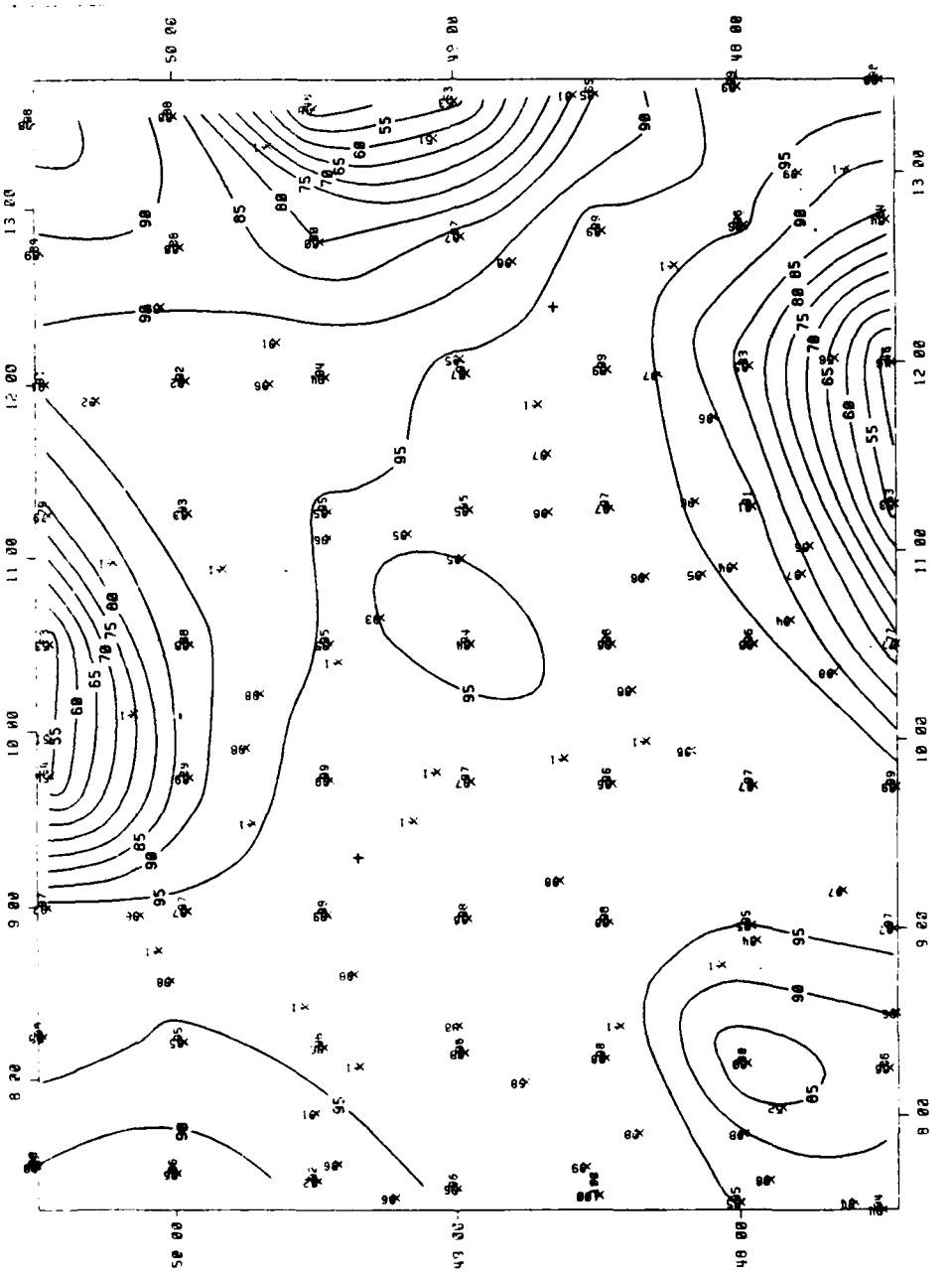


Figure A.28. Estimated Probability Fields, July, 02-04z, Ceiling .GT. 1000FT AGL. (26 missing observations.)

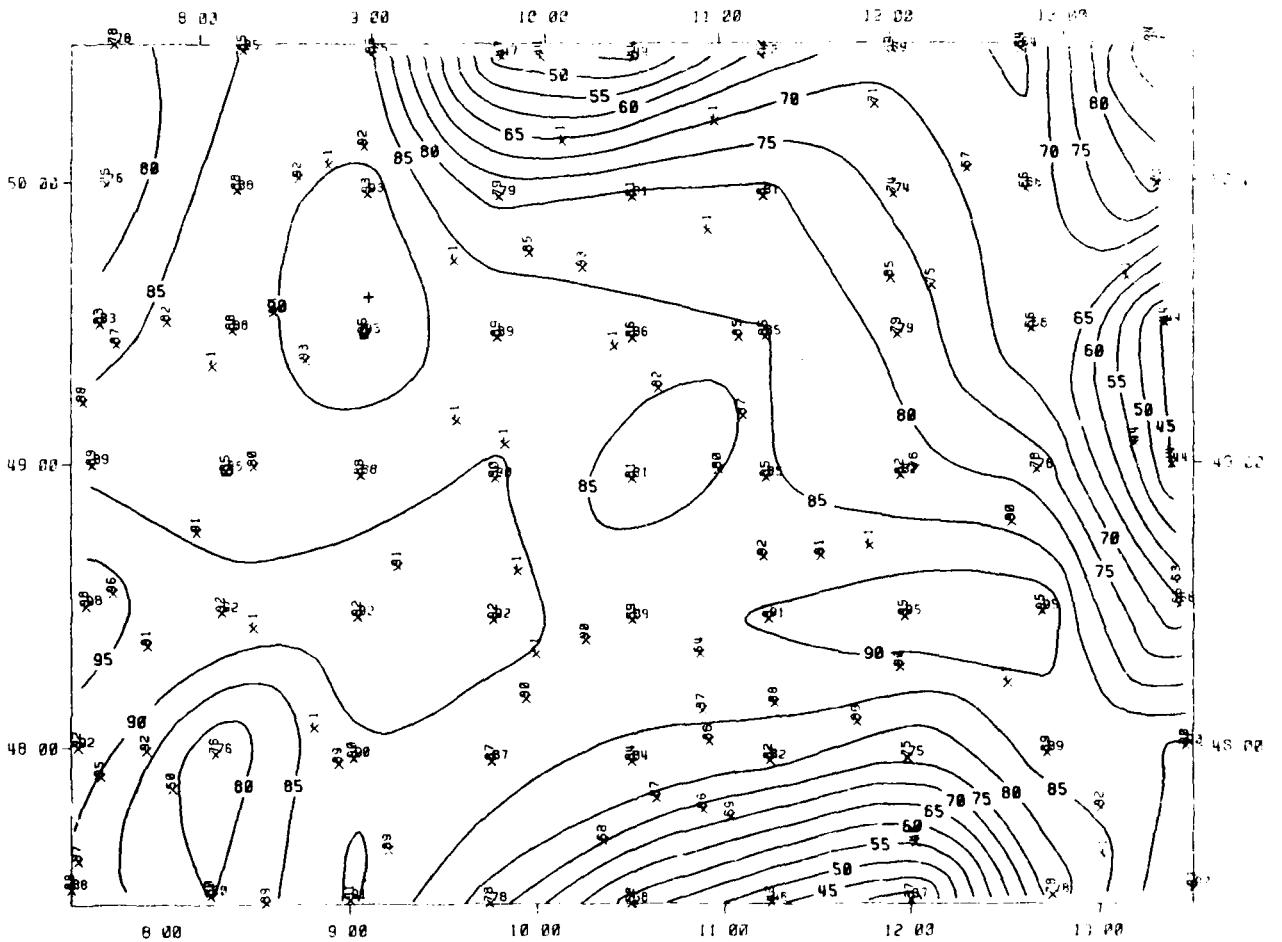


Figure A.29. Estimated Probability Fields, July, 02-04Z, Ceiling .GT. 3000FT AGL. (26 missing observations.)

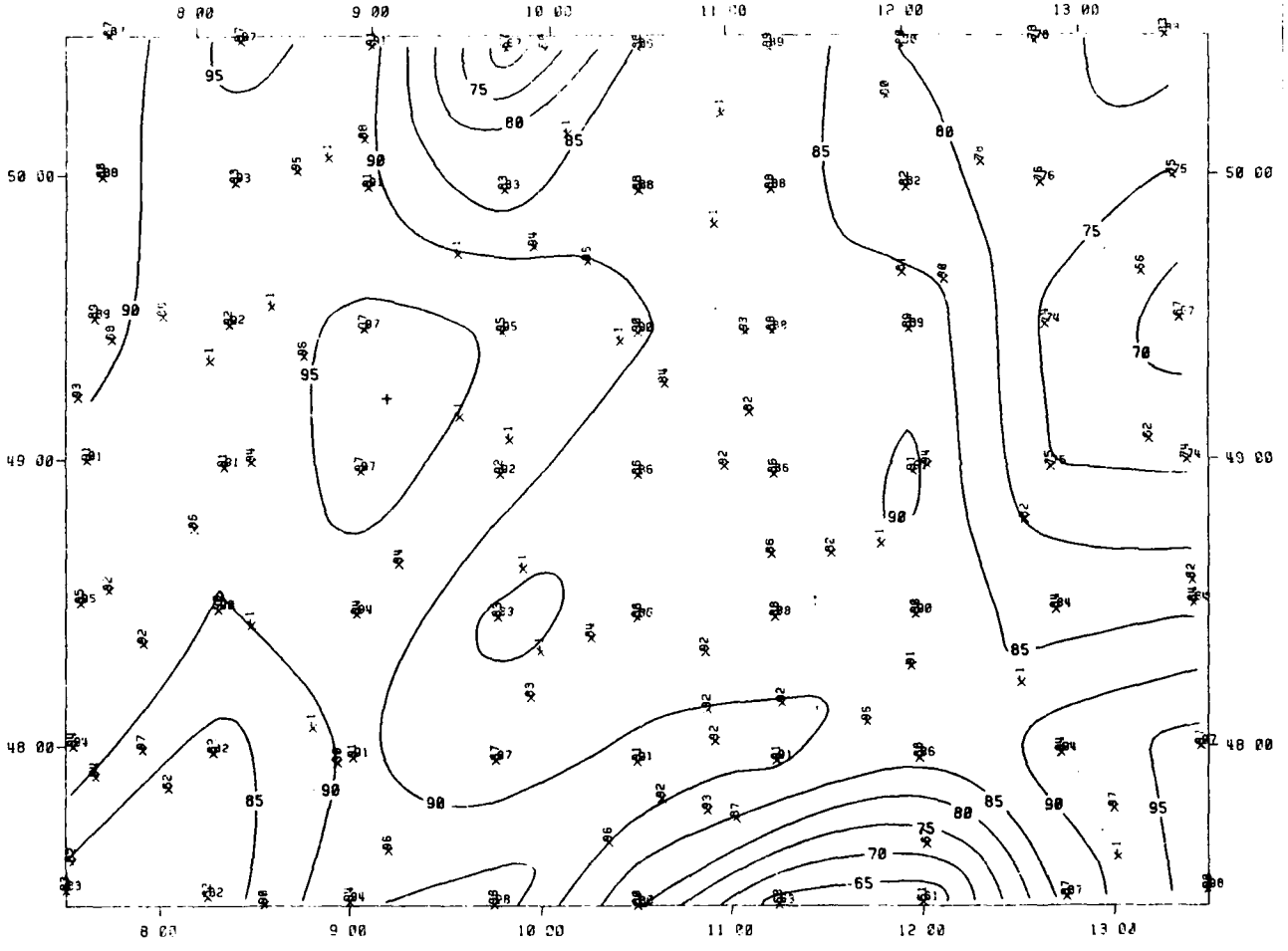


Figure A.30. Estimated Probability Fields, July, 02-04Z, Visibility .GT. 2MI. (22 missing observations.)

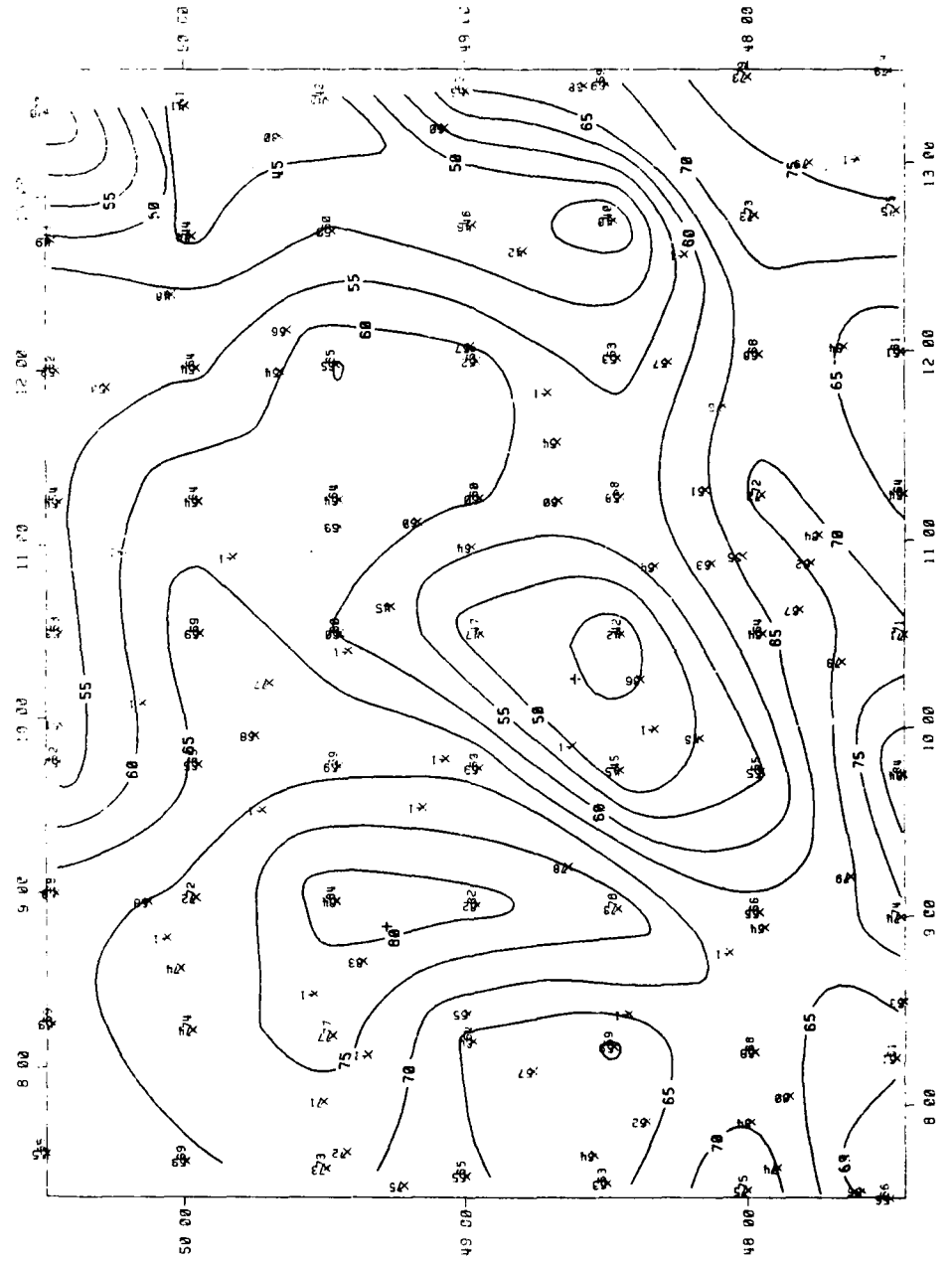


Figure A.31. Estimated Probability Fields, July, 02-04Z, Visibility GT. 5MI. (22 missing observations.)

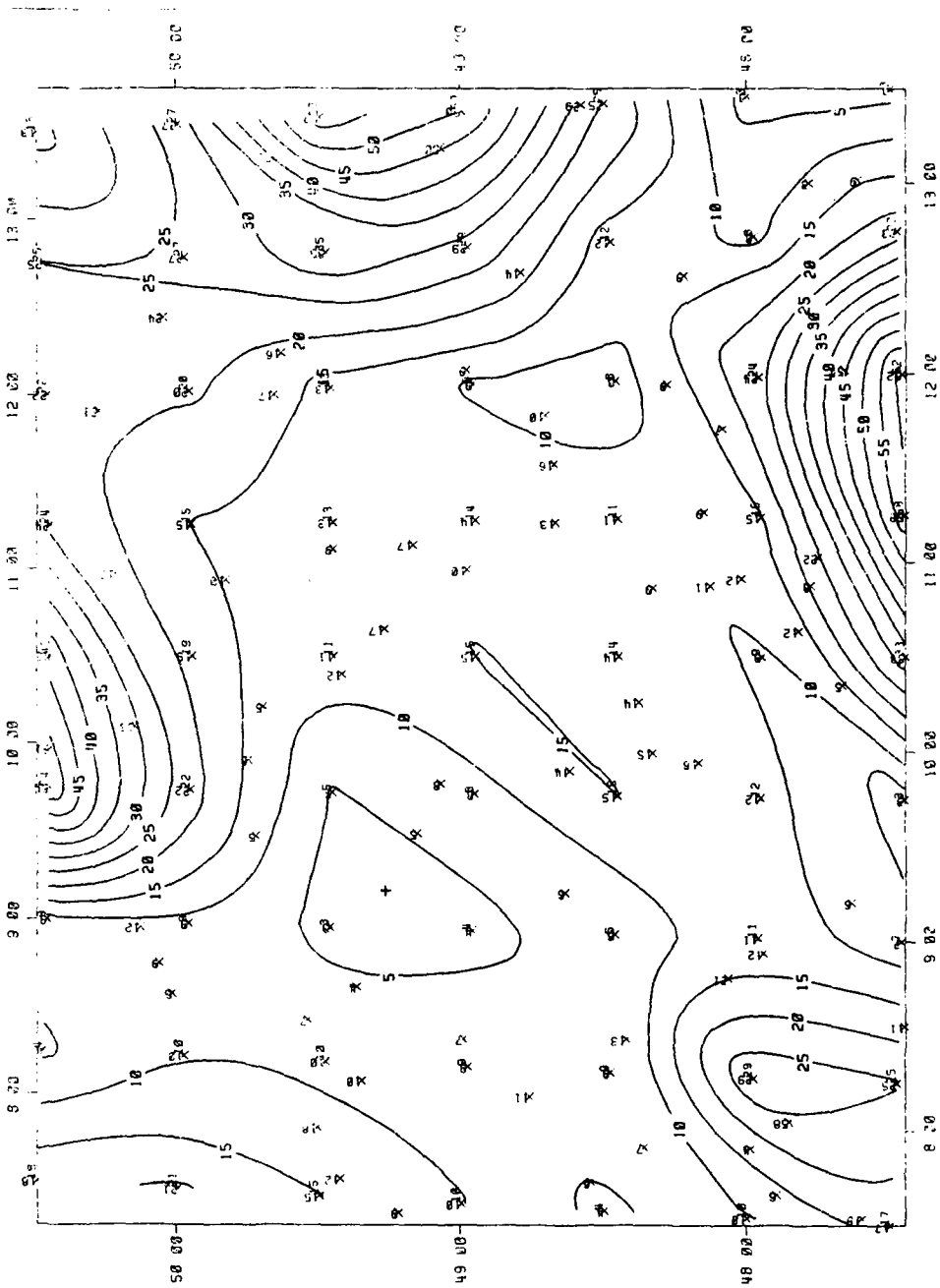


Figure A.32. Estimated Probability Fields, July, 02-04z, Joint CIG/VIS .LT. 1000/2.

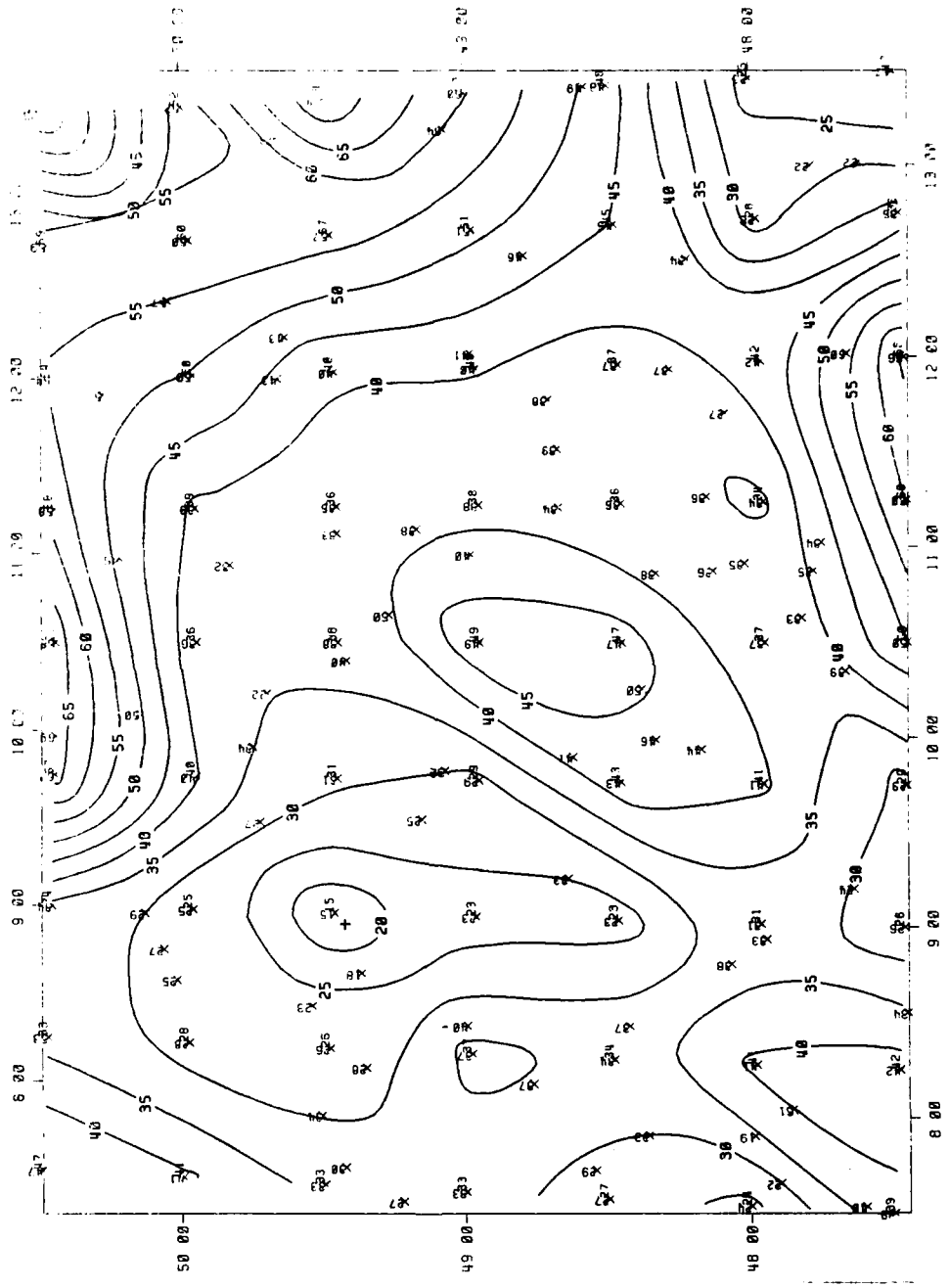


Figure A.33. Estimated Probability Fields, July, 02-04z, Joint CIG/VIS .LT. 3000/5.

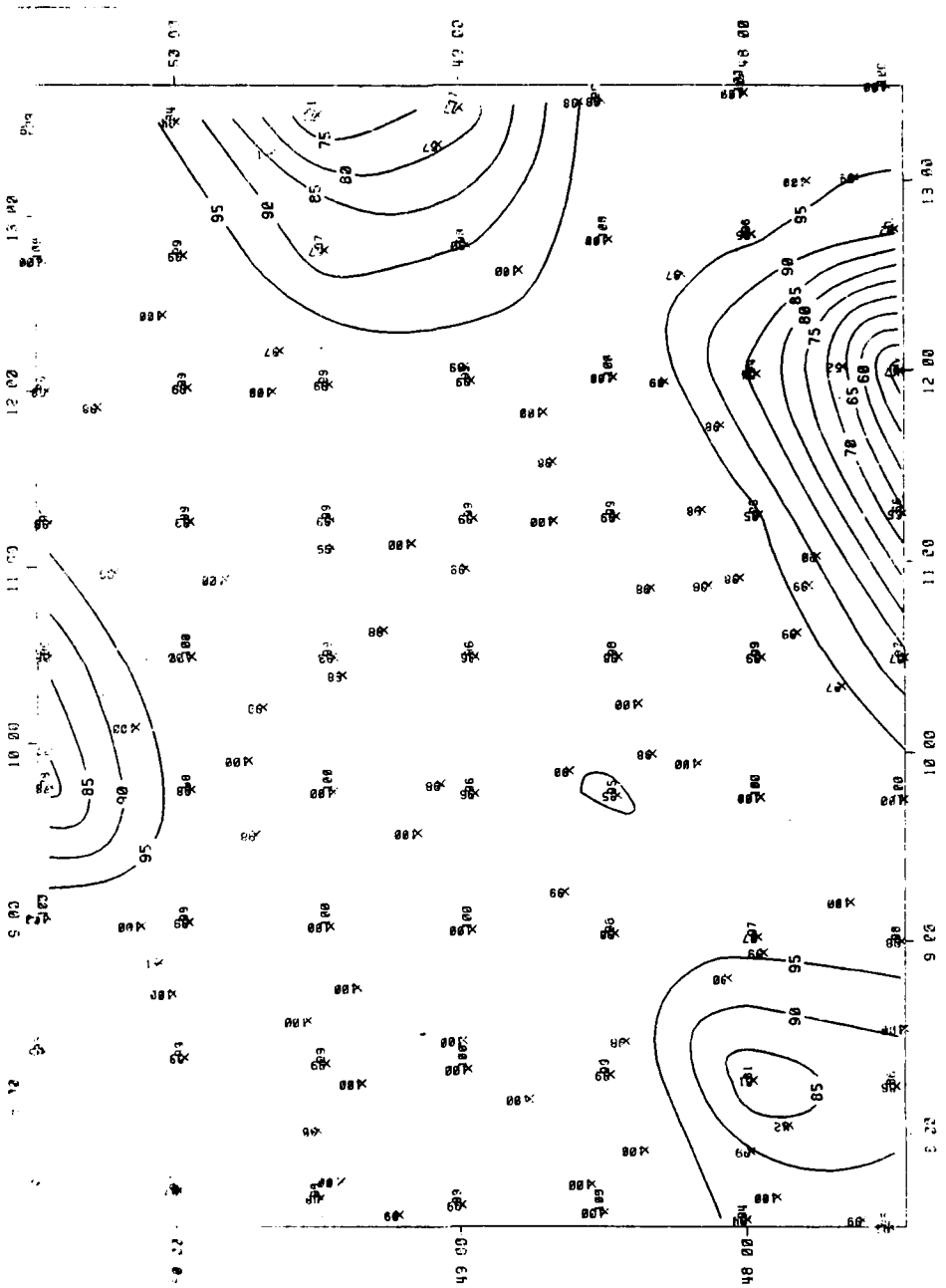


Figure A.34. Estimated Probability Fields,
July, 14-16Z, Ceiling .GT. 1000FT AGL.
(5 missing observations.)

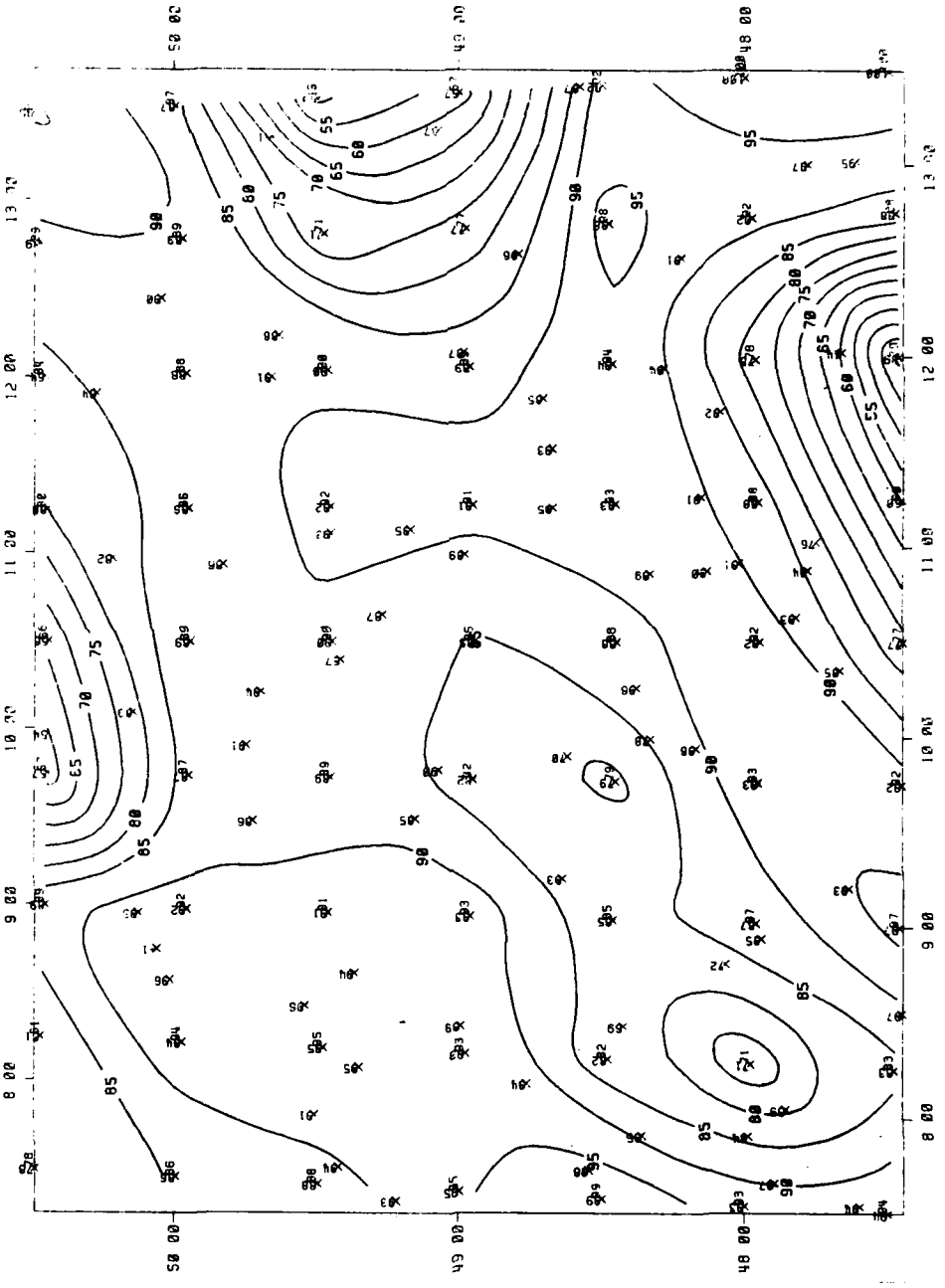


Figure A.35. Estimated Probability Fields, July, 14-16Z, Ceiling .GT. 3000FT AGL. (5 missing observations.)

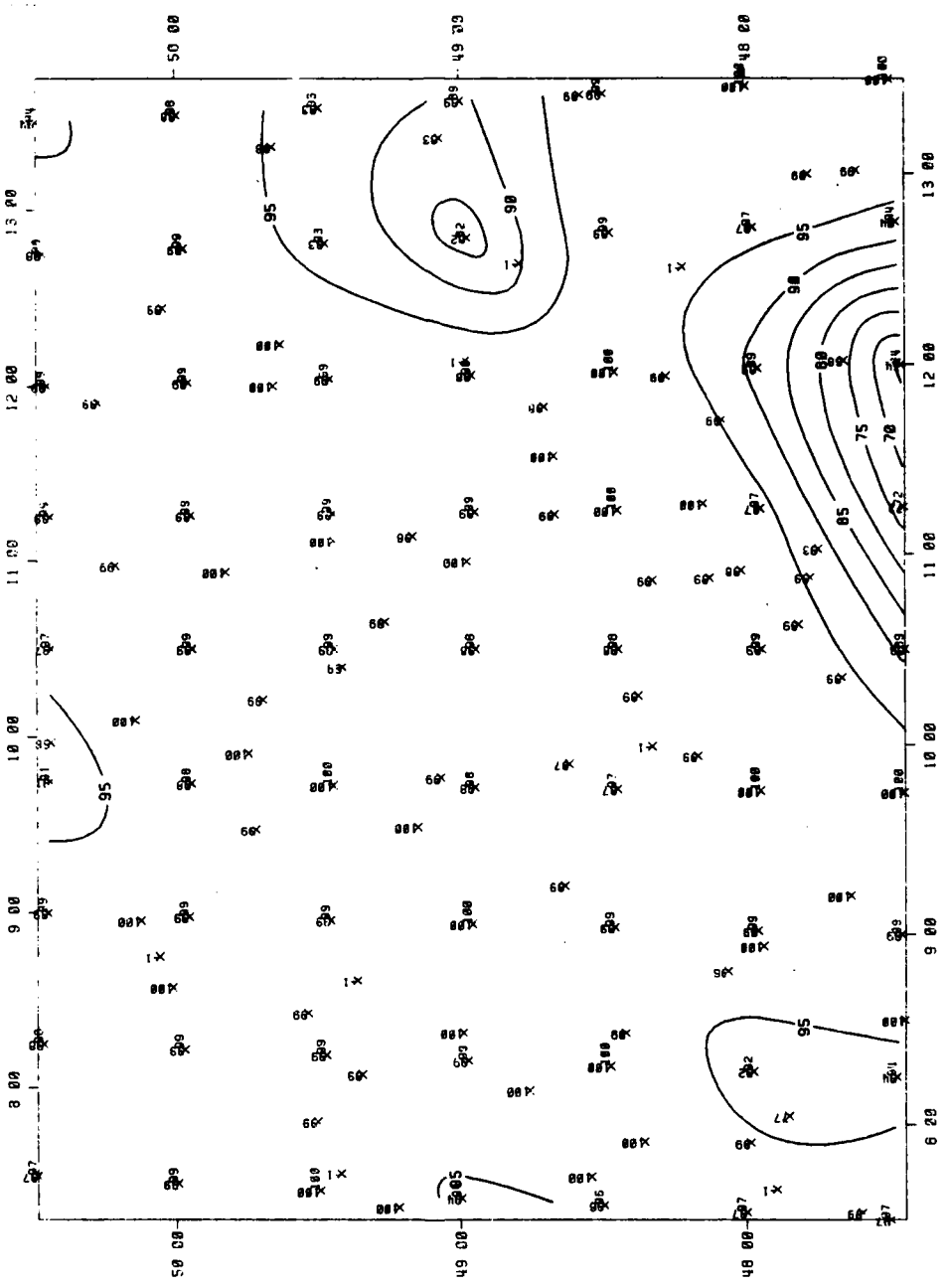


Figure A.36. Estimated Probability Fields, July, 14-16Z, Visibility .GT. 2MI. (8 missing observations.)

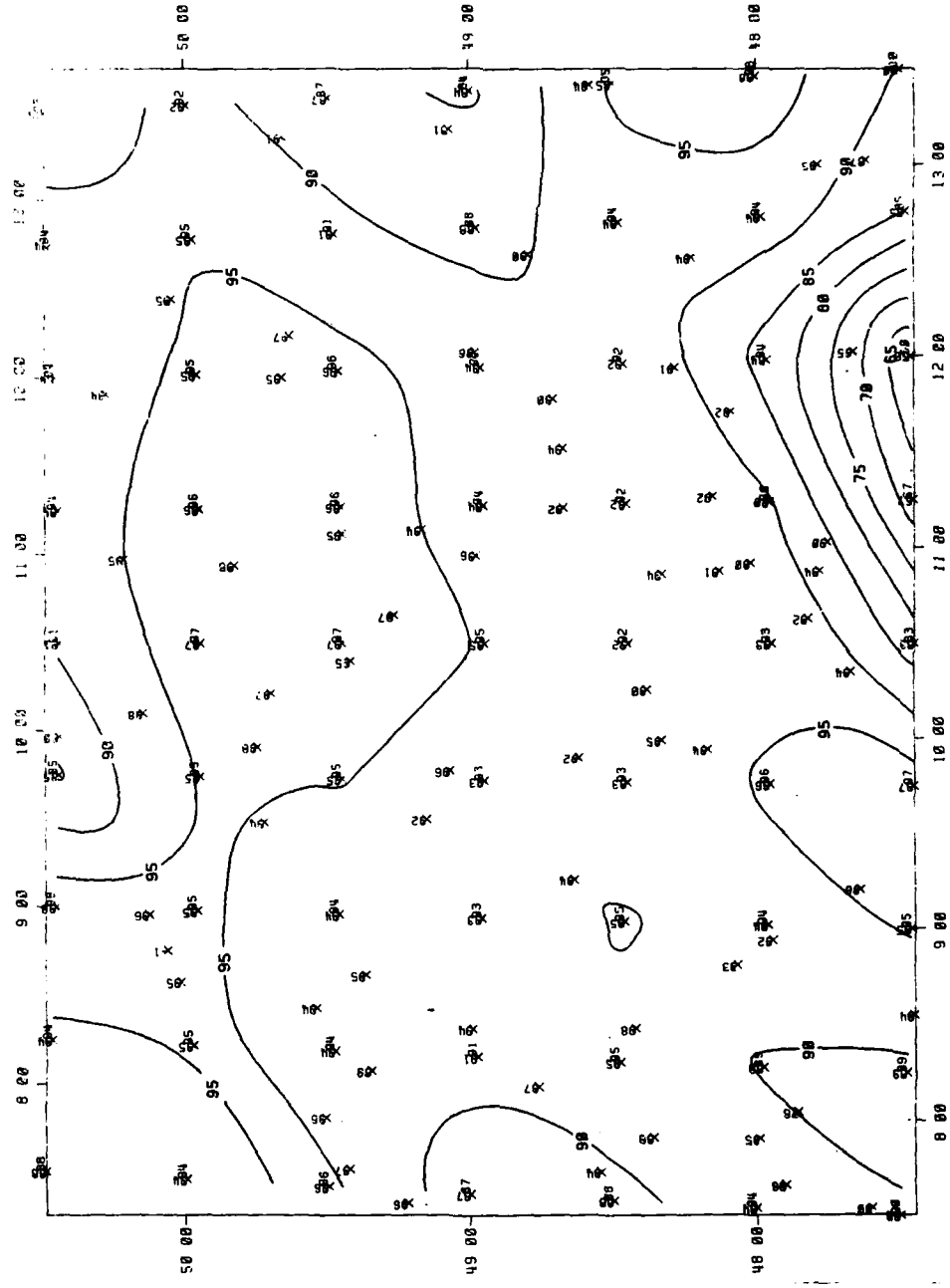


Figure A.37. Estimated Probability Fields, July, 14-16Z, Visibility .GT. 5MI. (1 missing observation.)

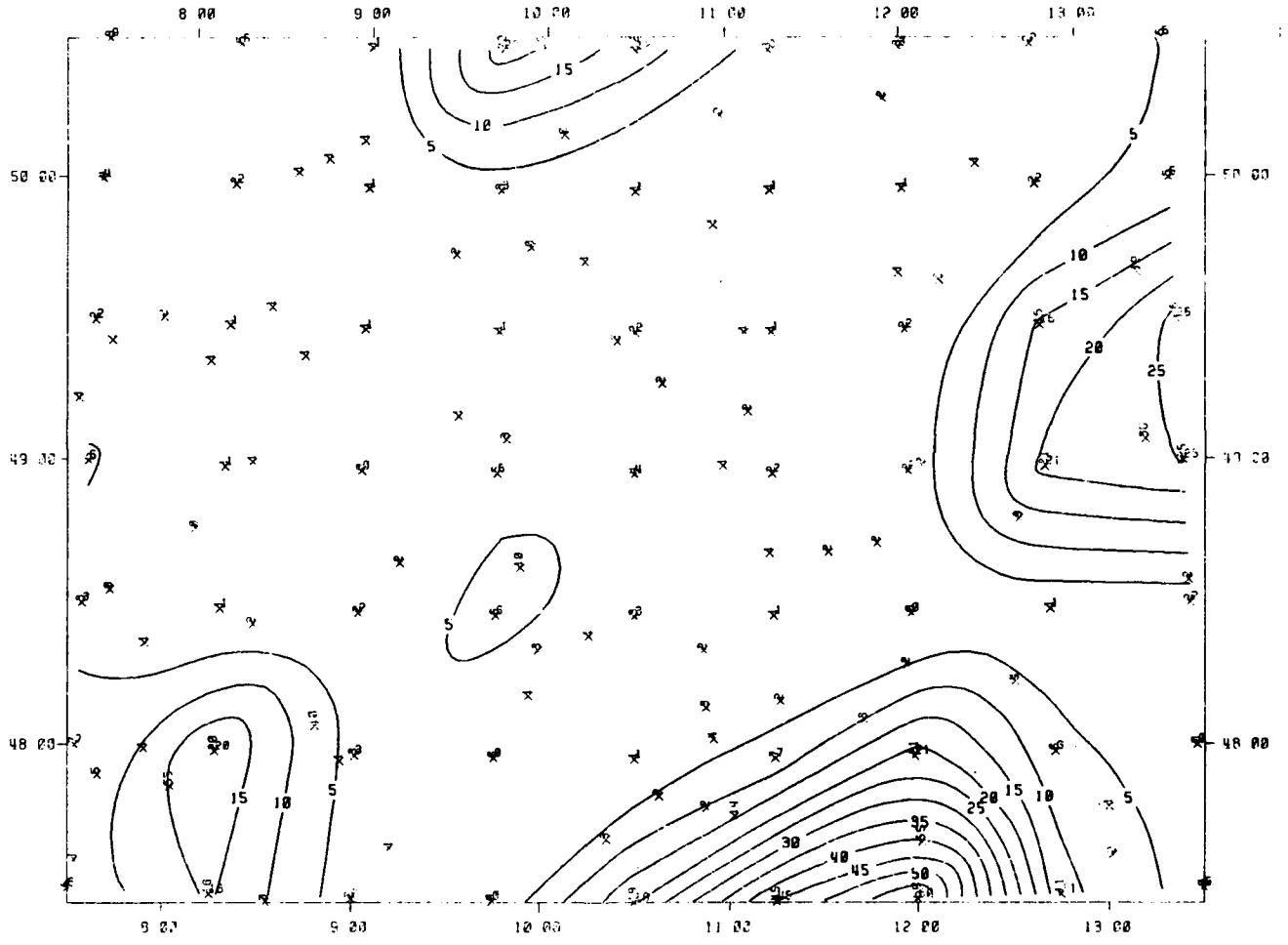


Figure A.38. Estimated Probability Fields,
July, 14-16Z, Joint CIG/VIS .LT. 1000/2.

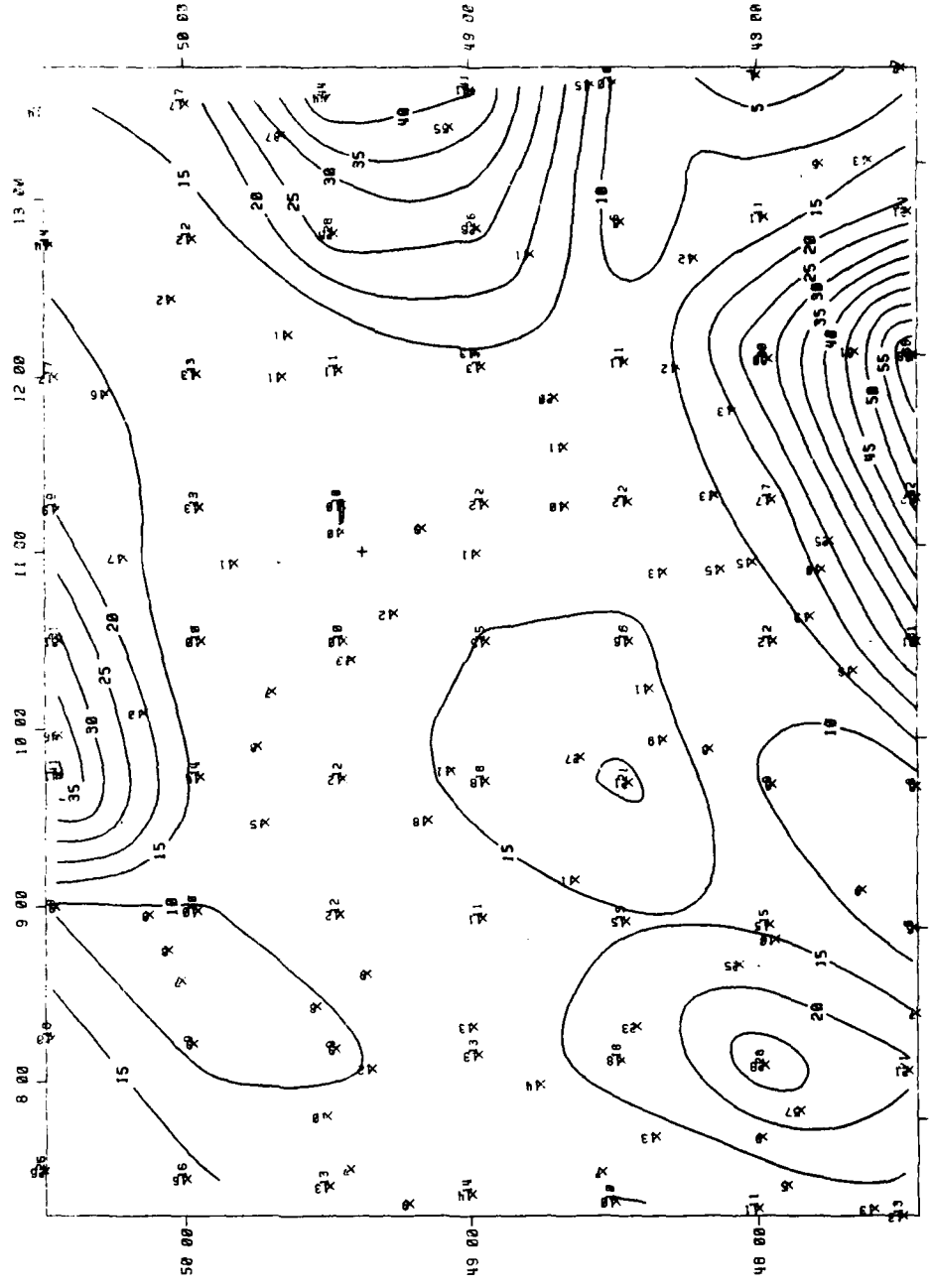


Figure A.39. Estimated Probability Fields, July, 14-16z, Joint CIG/VIS .LT. 3000/5.

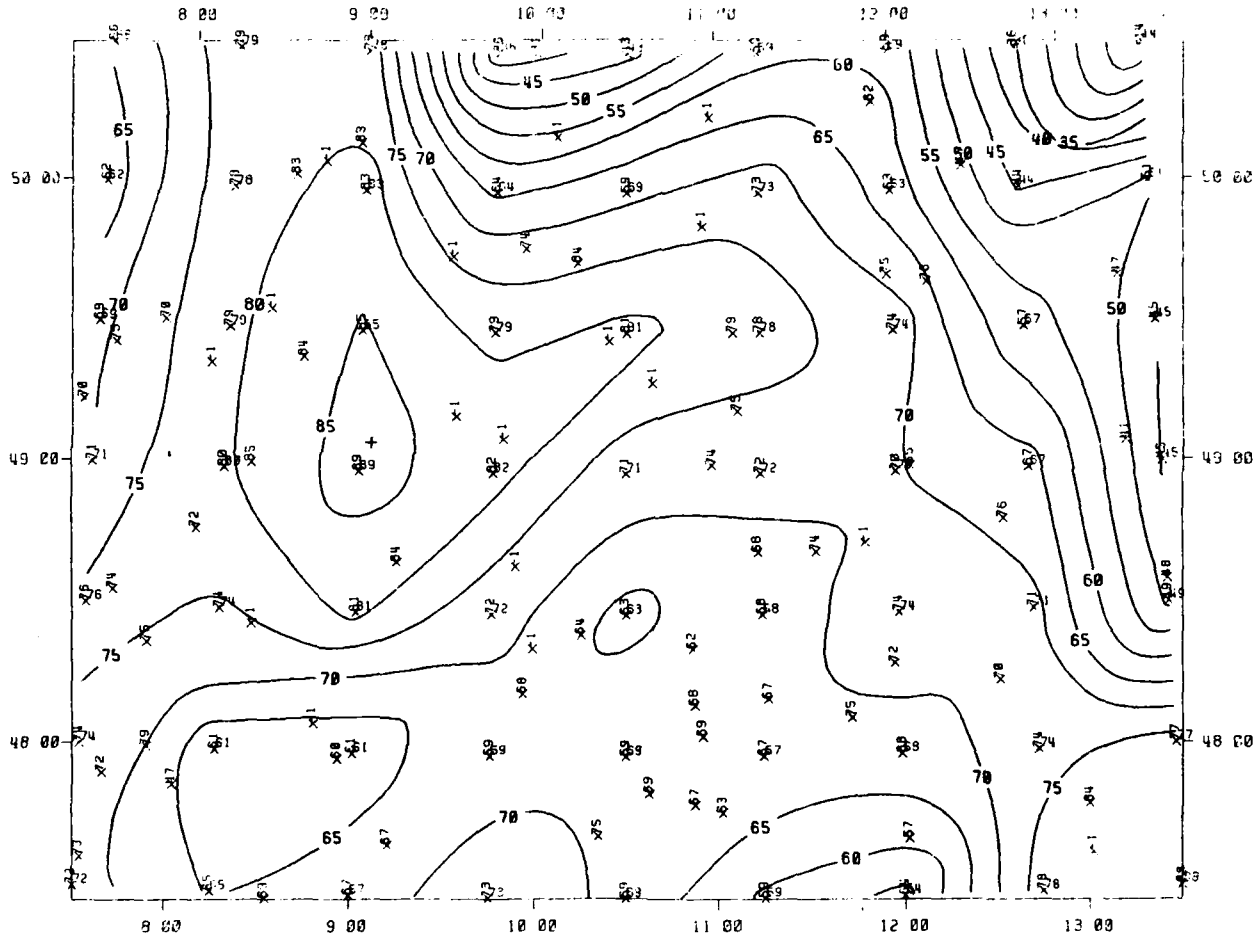


Figure A.40. Estimated Probability Fields, October, 02-04Z, Ceiling .GT. 1000FT AGL. (22 missing observations.)

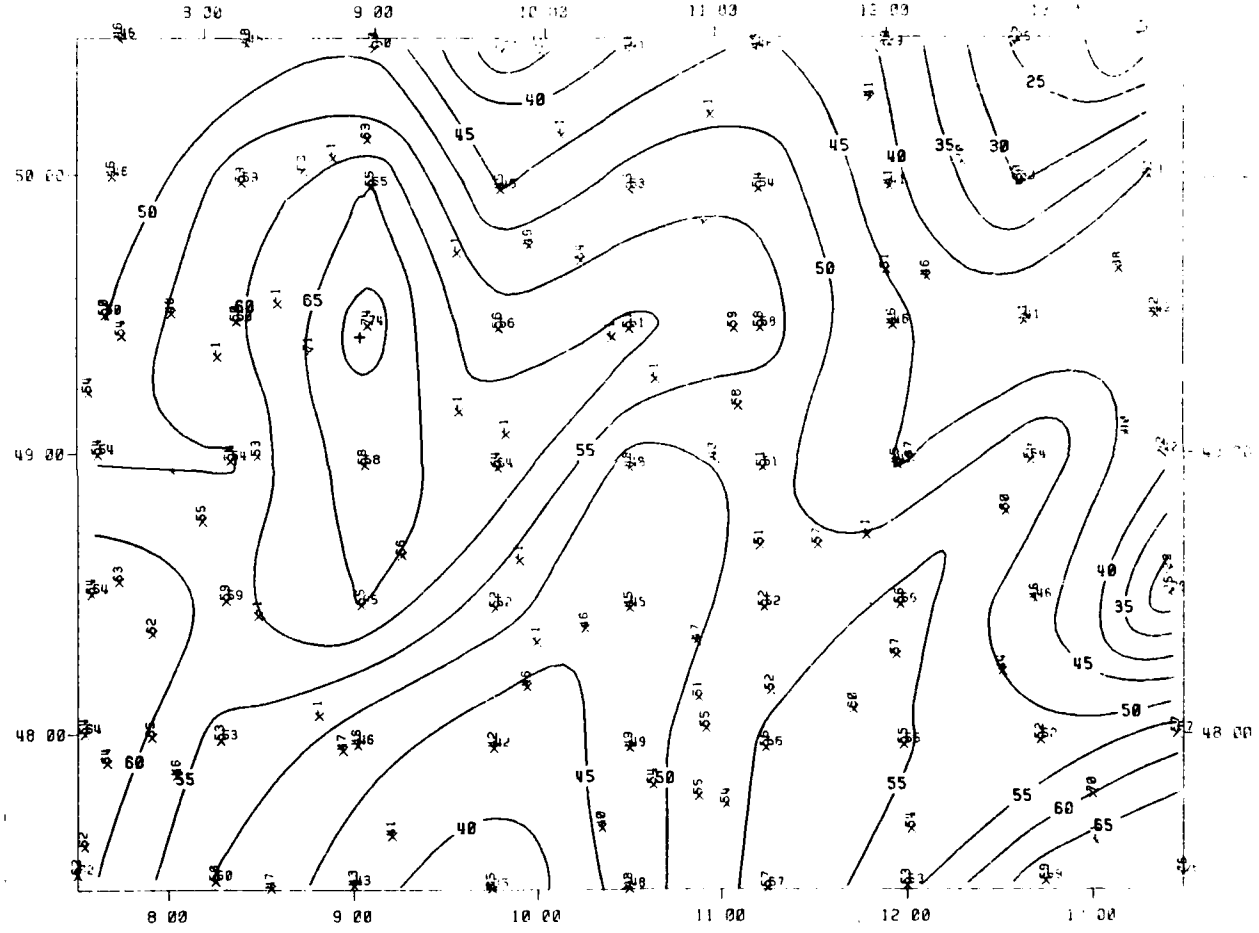


Figure A.41. Estimated Probability Fields, October, 02-04Z, Ceiling .GT. 3000FT AGL. (22 missing observations.)

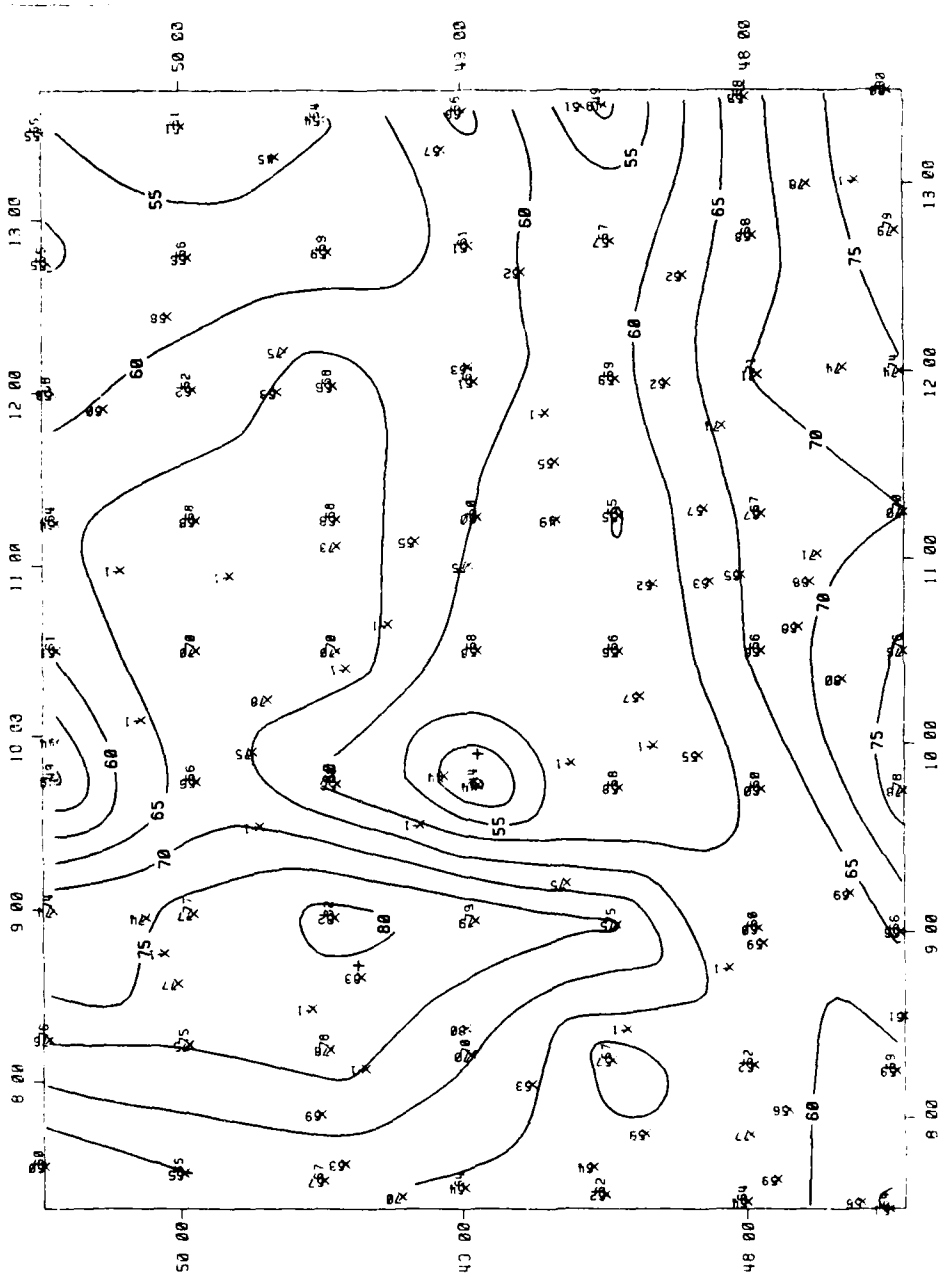


Figure A.42. Estimated Probability Fields, October, 02-04Z, Visibility .GT. 2MI. (21 missing observations.)

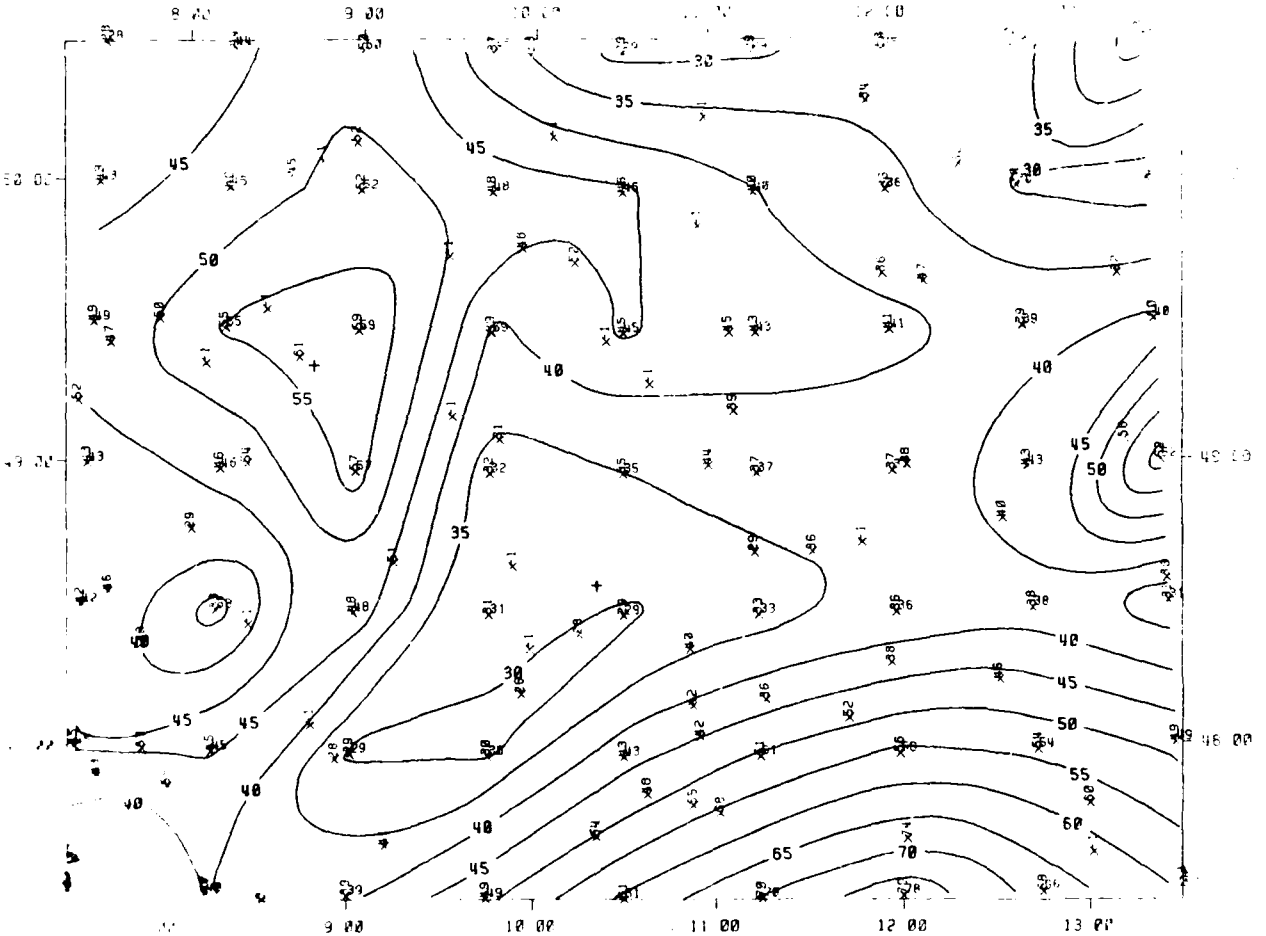


Figure A.43. Estimated Probability Fields,
October, 02-04Z, Visibility .GT. 5MI.
(1 missing observations.)

AD-A118 429 AIR FORCE ENVIRONMENTAL TECHNICAL APPLICATIONS CENTER--ETC F/G 9/2
OBJECTIVE ANALYSIS OF CLIMATOLOGICAL PROBABILITY DATA.(U)
JUL 82 B E LILIUS, F C WIRSING, R M COX
UNCLASSIFIED USAFETAC/TN-82/003 SRI-AR-FR50 1A1 NL

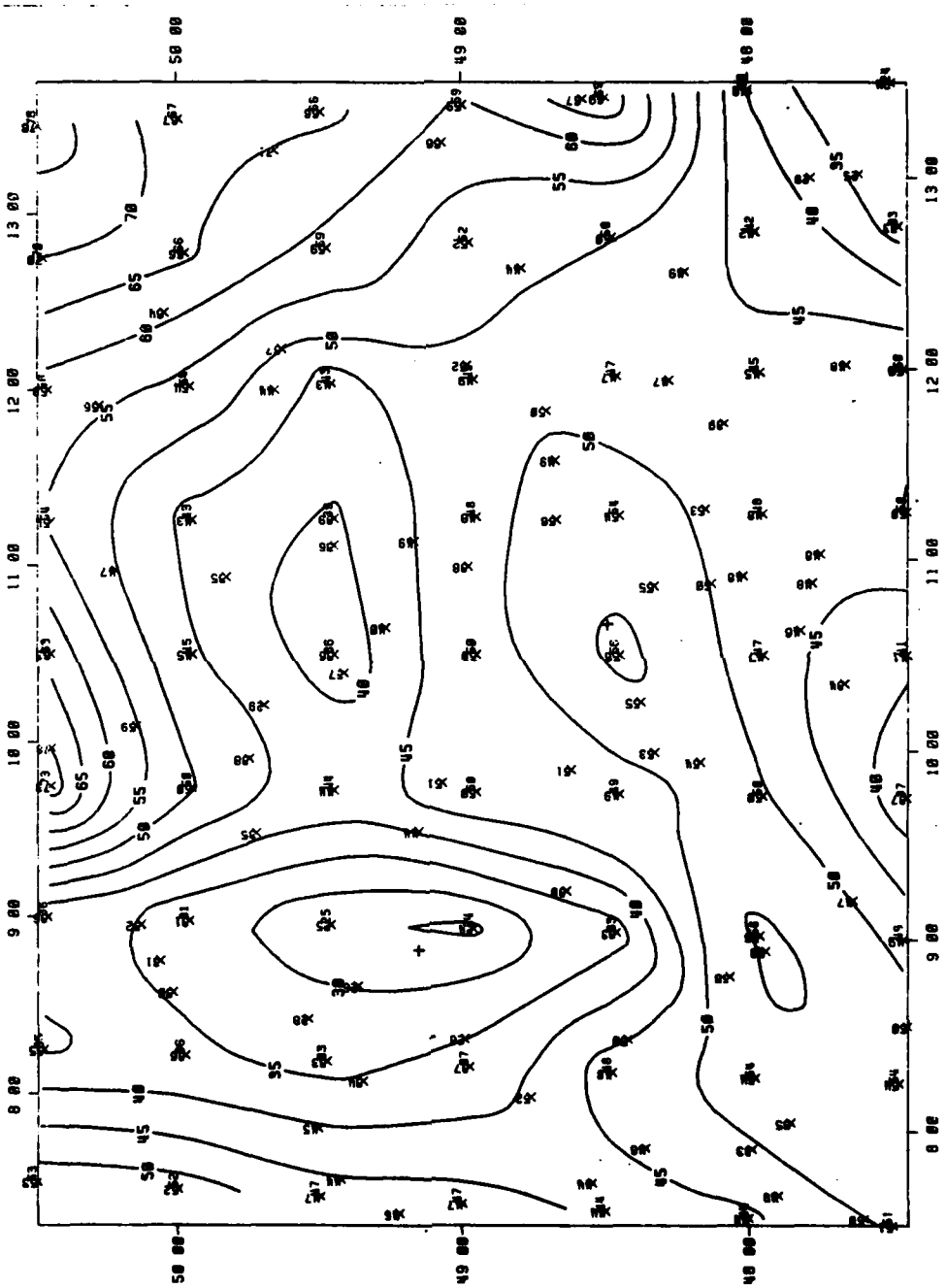


Figure A.44. Estimated Probability Fields, October, 02-04Z, Joint CIG/VIS .LT. 1000/2.

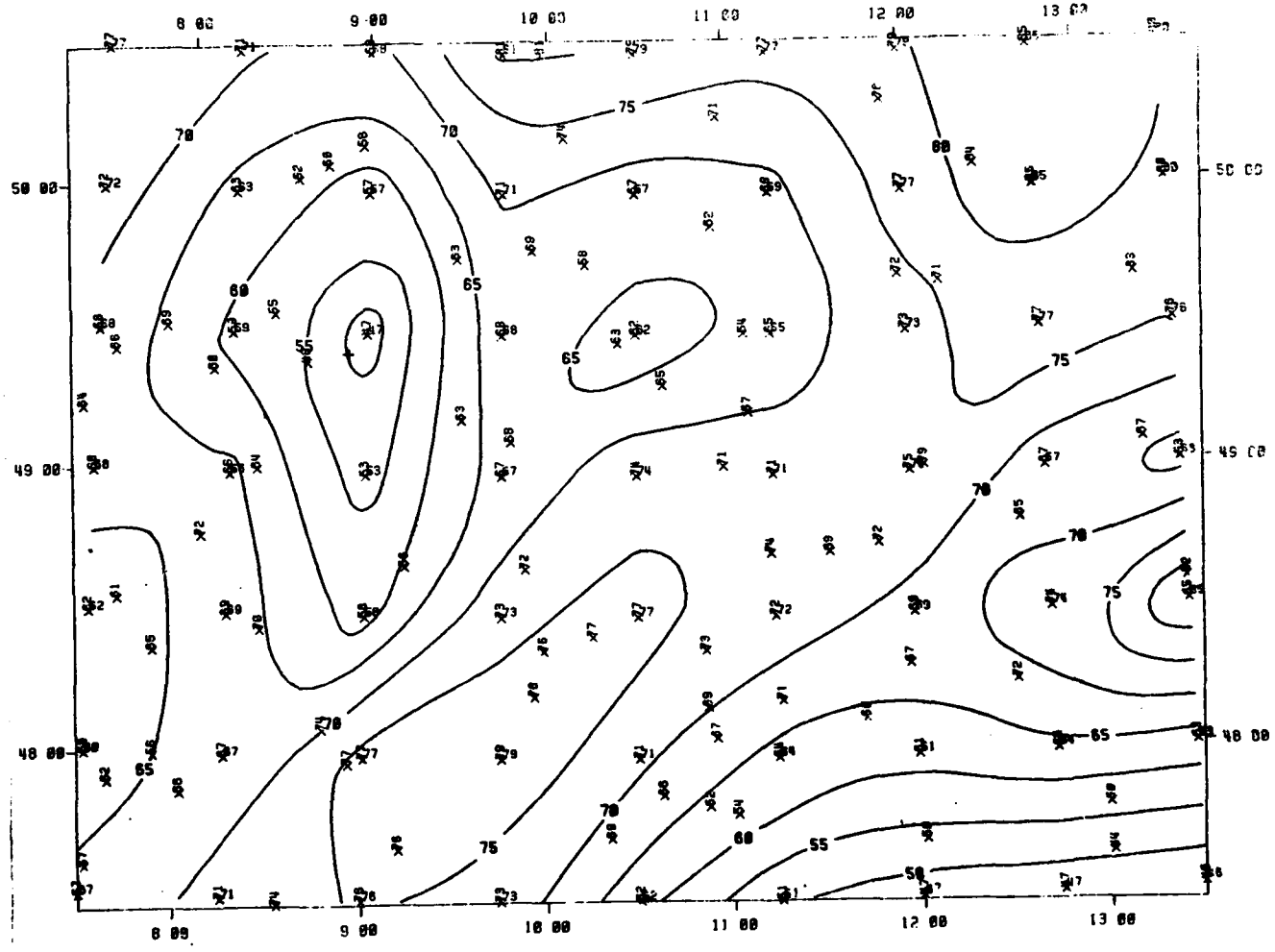


Figure A.45. Estimated Probability Fields, October, 02-04Z, Joint CIG/VIS .LT. 3000/5.

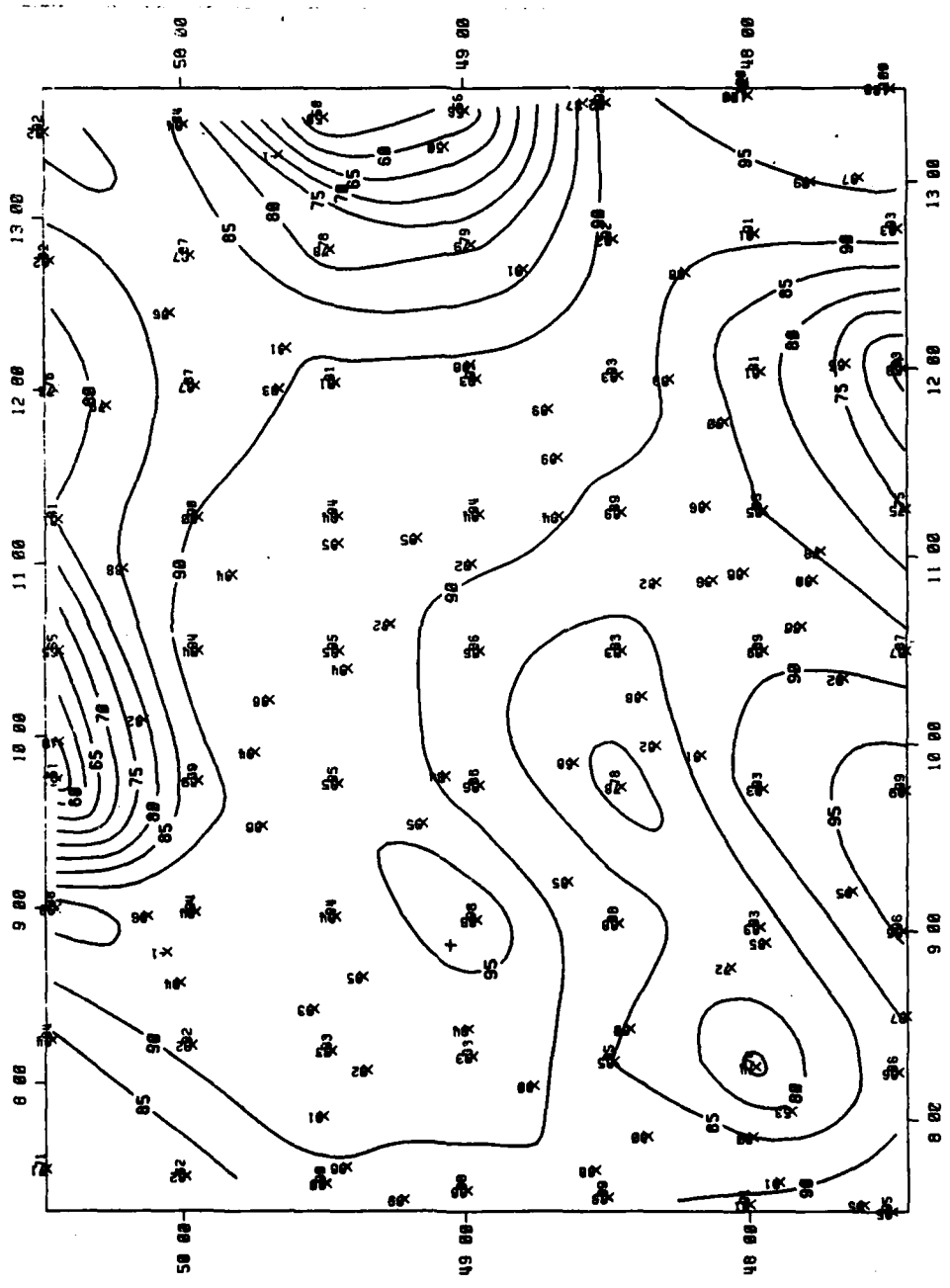


Figure 46. Estimated Probability Fields,
October, 14-16Z, Ceiling .GT. 1000FT AGL.
(4 missing observations.)

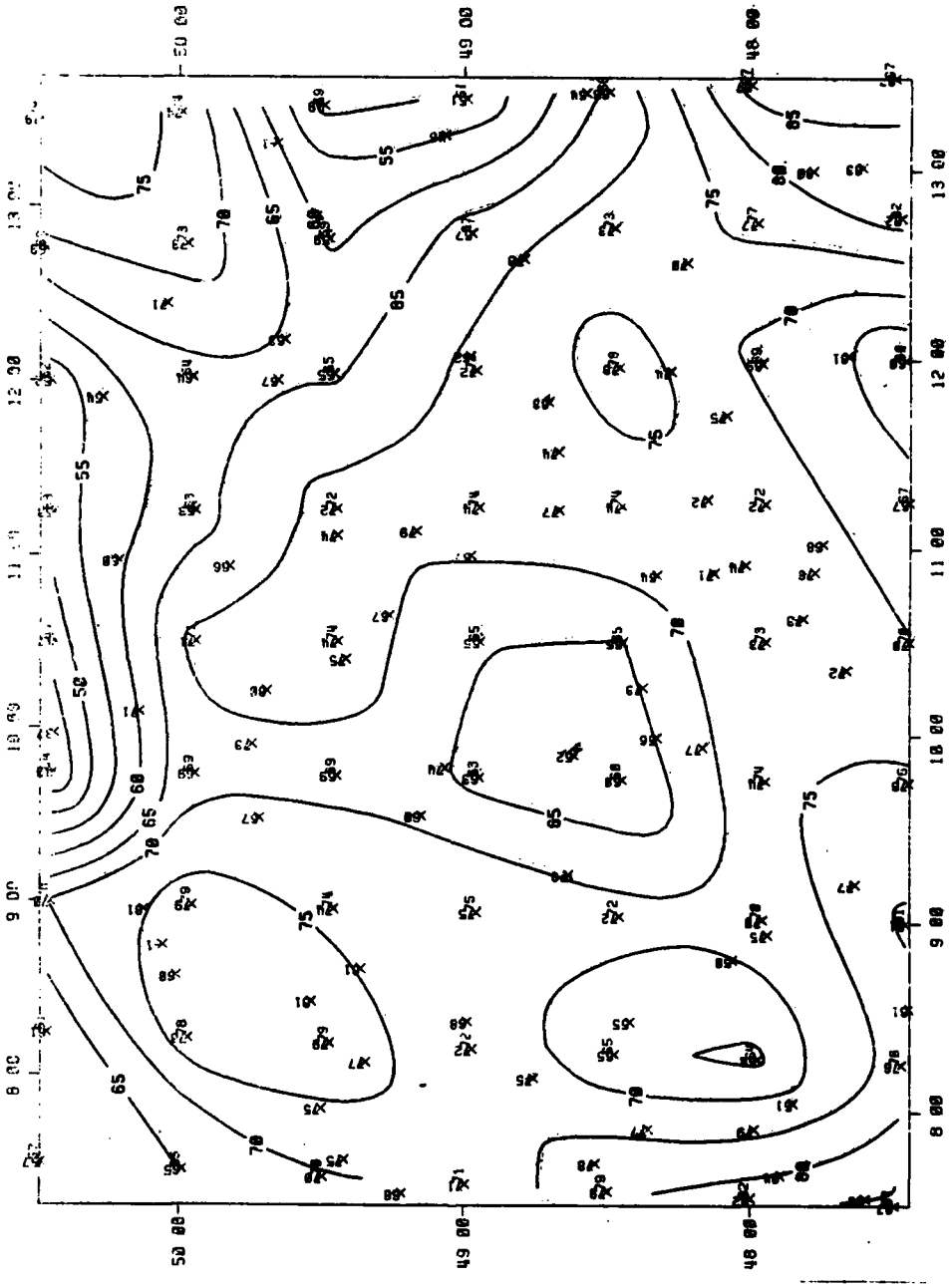


Figure A.47. Estimated Probability Fields, October, 14-16Z, Ceiling .GT. 3000FT AGL. (4 missing observations.)

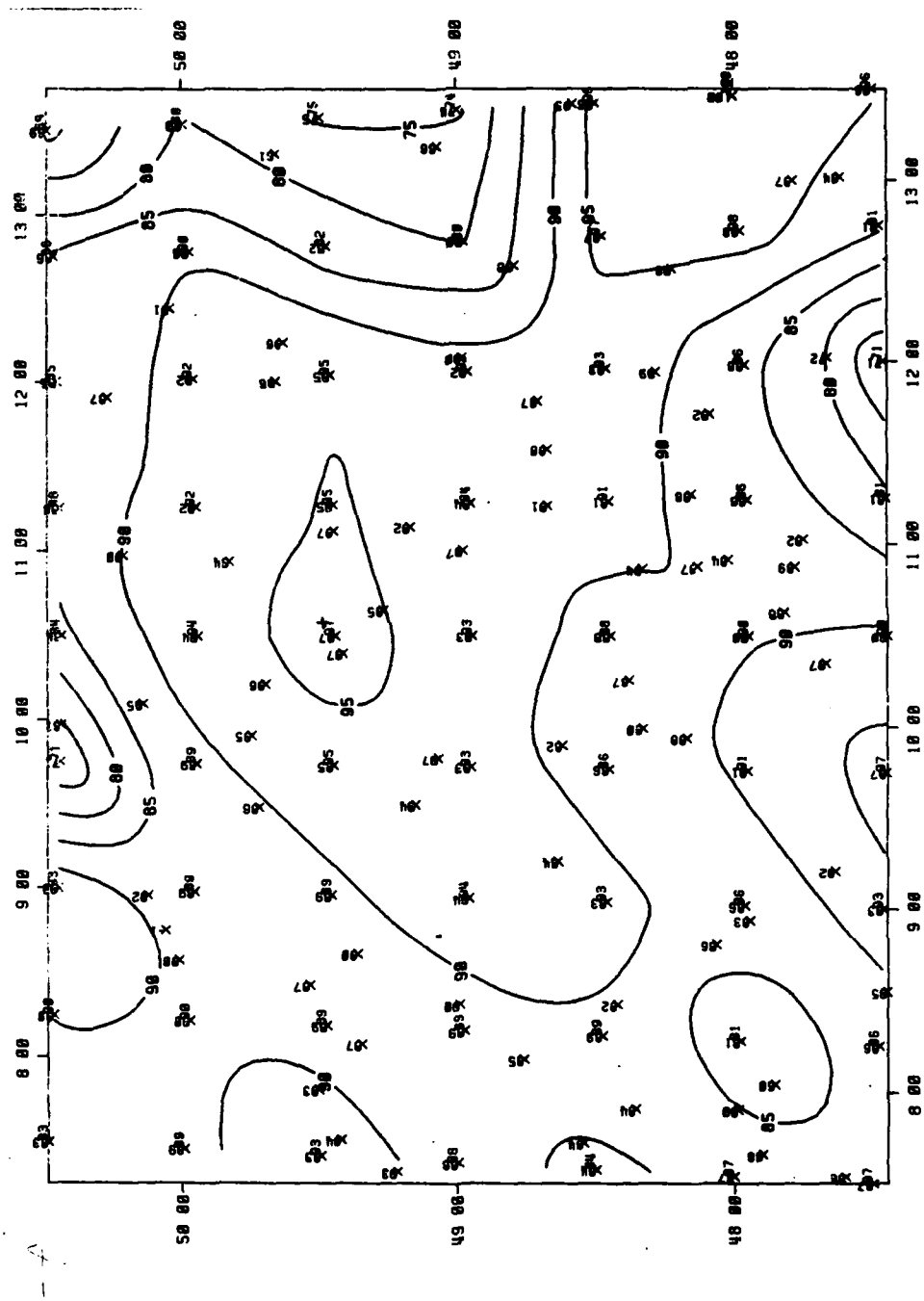


Figure A.48. Estimated Probability Fields, October, 14-16Z, Visibility .GT. 2MI. (1 missing observation.)

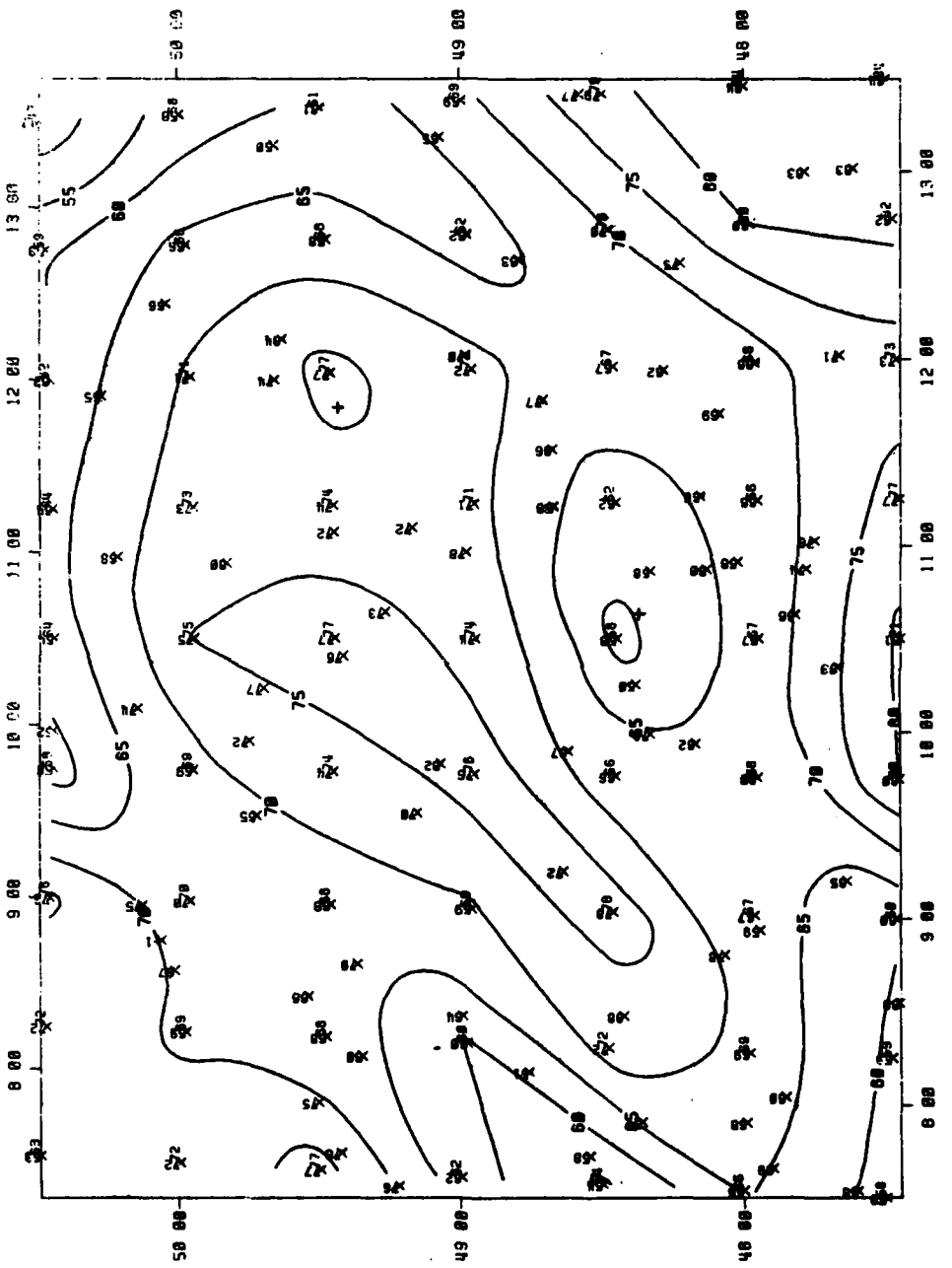


Figure A.49. Estimated Probability Fields, October, 14-16Z, Visibility .GT. 5MI. (1 missing observation.)

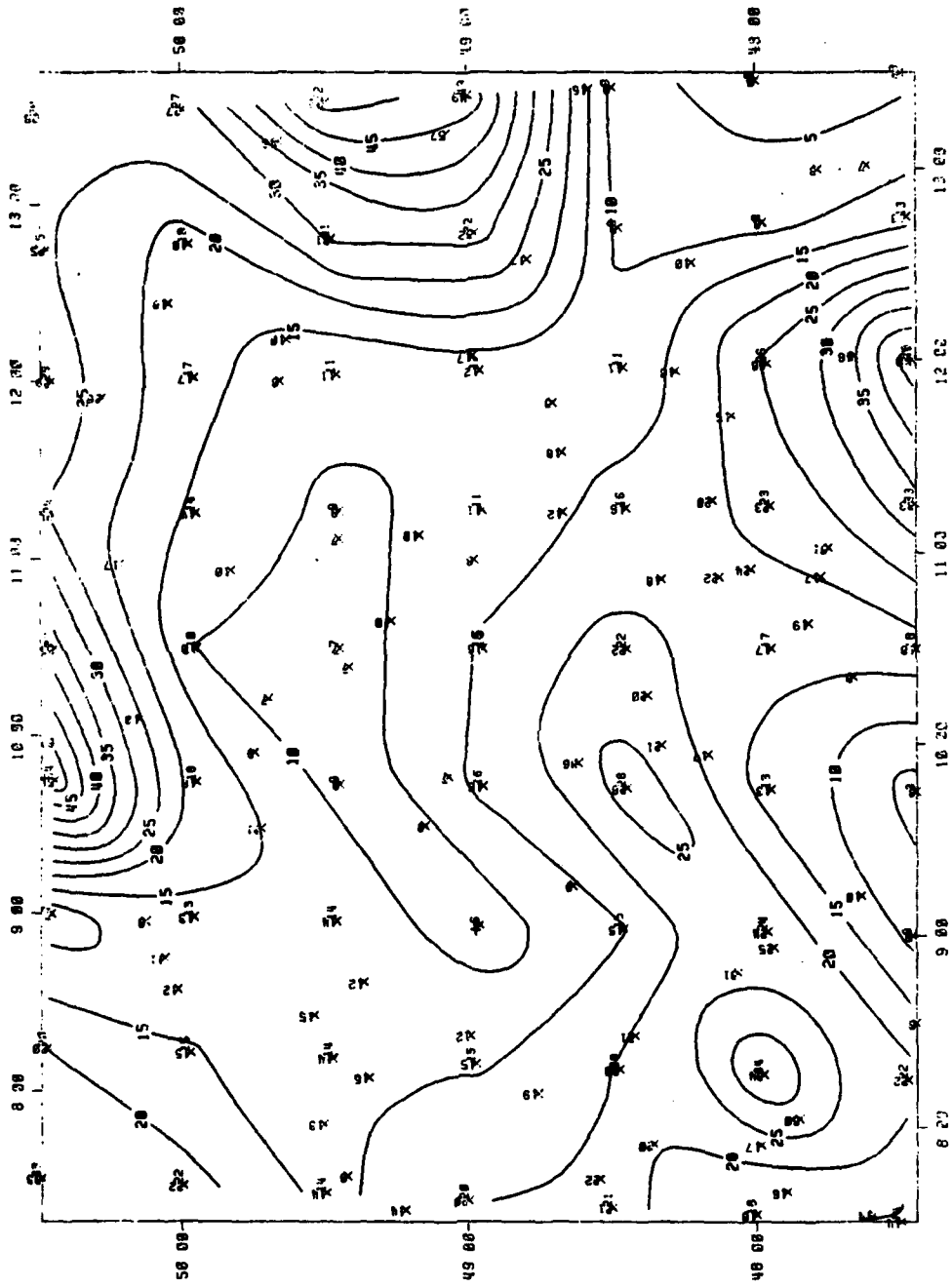


Figure A.50. Estimated Probability Fields, October, 14-16Z, Joint CIG/VIS .LT. 1000/2.

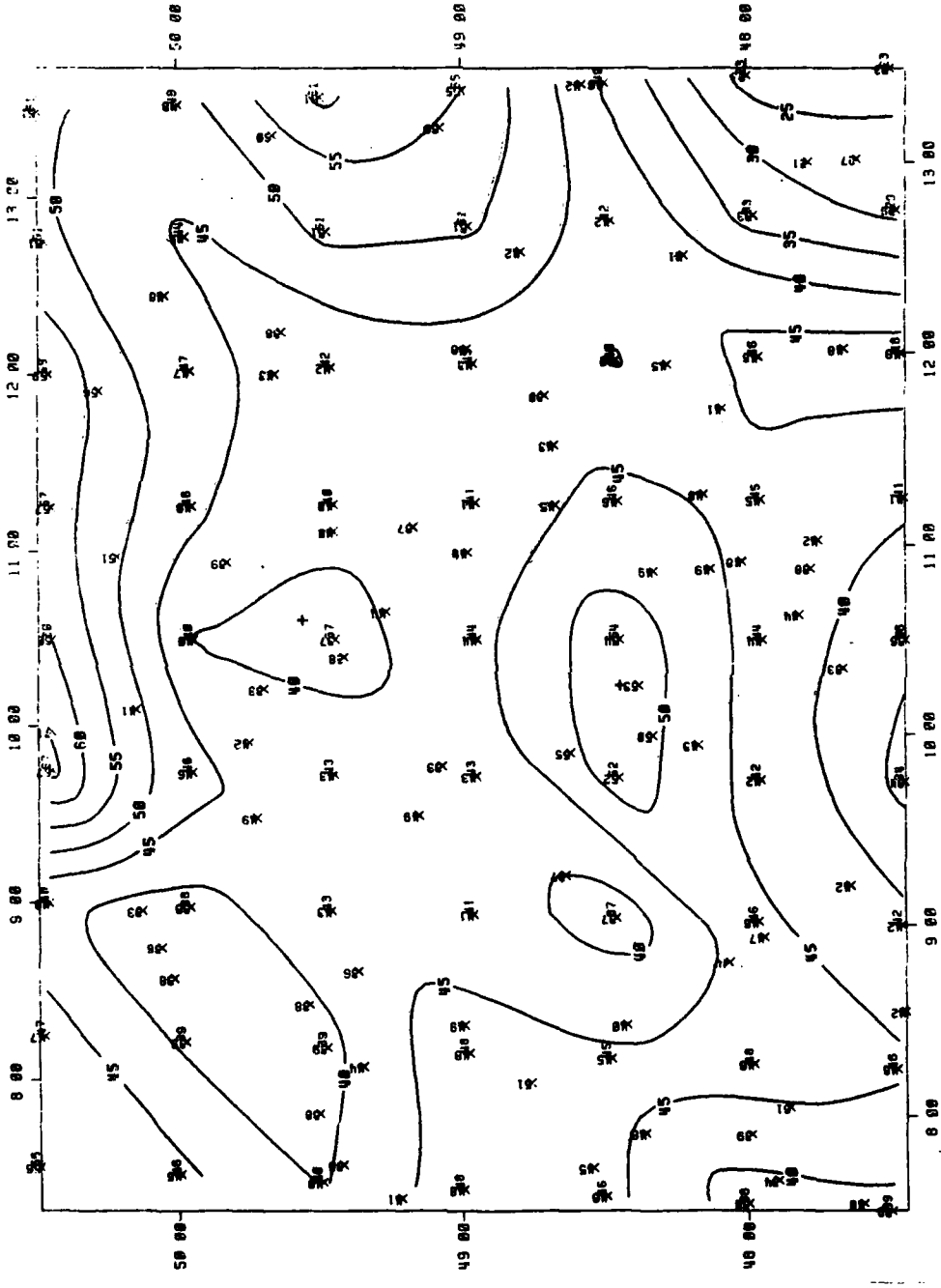


Figure A.51. Estimated Probability Fields, October, 14-16Z, Joint CIG/VIS .LT. 3000/5.

GLOSSARY

AFGWC	Air Force Global Weather Central
AGL	Above ground level
AWS	Air Weather Service
CDF	Cumulative Distribution Function
CIG	Ceiling
CPU	Central Processing Unit
DAR	Data Automation Request
DATSAV	Data Save
DN	(USAFETAC) Aerospace Sciences Branch
DND	(USAFETAC) Data Base Development Section
DNO	(USAFETAC) Operations Applications Development Section
.GT.	Greater than
LSS	Line-segment-selection
.LT.	Less than
MOS	Model Output Statistics
POR	Period of Record
RMS	Root-mean-square
RMSE	Root-mean-square error
RUSSWO	Revised Uniform Summary of Surface Weather Observations
USAFETAC	United States Air Force Environmental Technical Applications Center
VIS	Visibility
WMO	World Meteorological Organization
WWMCCS	Worldwide Military Command and Control System
3DNEPH	Three-Dimensional Nephanalysis